HAL
open science

Bayesian Approach to Action Selection and Attention Focusing. An Application in Autonomous Robot Programming.

Carla Maria Chagas E Cavalcante Koike

To cite this version:

Carla Maria Chagas E Cavalcante Koike. Bayesian Approach to Action Selection and Attention Focusing. An Application in Autonomous Robot Programming.. Other [cs.OH]. Institut National Polytechnique de Grenoble - INPG, 2005. English. NNT : . tel-00011138

HAL Id: tel-00011138

https://theses.hal.science/tel-00011138v1

Submitted on 1 Dec 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

N^o attribué par la bibliothèque
|__/__/__/__/__/__/__/__/__|

THÈSE

pour obtenir le grade de

DOCTEUR DE L'INPG

Spécialité : **Imagerie, Vision, Robotique**

préparée au laboratoire GRAVIR et l'INRIA Rhône-Alpes, dans le cadre de *l'École Doctorale* **Mathématiques, Sciences et Technologies de l'Information, Informatique**

présentée et soutenue publiquement par

Carla Maria CHAGAS e CAVALCANTE KOIKE

le 14 novembre 2005

Titre :

Bayesian Approach to Action Selection and Attention Focusing
An Application in Autonomous Robot Programming

Co-directeurs de Thèse : Emmanuel MAZER et Pierre BESSIÈRE

Composition du jury :

M.	Jean-Luc Schwartz	Président
Mme.	Leslie P. Kaelbling	Rapporteur
M.	Jacques Droulez	Rapporteur
M.	Guilherme C. Carvalho	Examinateur
M.	Emmanuel Mazer	Directeur de thèse
M.	Pierre Bessière	Co-directeur de thèse

Abstract

Autonomous sensory-motor systems, situated in dynamic environments, must continuously answer the ultimate question: how to control motor commands knowing sensory inputs?

Solving this question is a very complex problem, because a huge flow of information must be treated under several restrictions: real-time constraints, bounded memory space, and limited processing power. One additional and major challenge is to deal with incomplete and imprecise information, usually occurring in dynamic environments.

In this thesis, we address the problem of controlling autonomous sensory-motor systems and propose a succession of cumulative hypotheses and simplifications. They are defined within a precise and strict mathematical framework, called Bayesian programming, an extension of Bayesian networks. This succession consists of five stages:

- Internal states summarise the sensory-motor situation to simplify modelling and break the exponential degradation in performance because of dimensionality;
- The first-order Markov assumption, stationarity and Bayes filters reduce time dependence without neglecting the influence of the past;
- Partial independence between different domains of interest can be exploited to reduce further the dimensionality of the problem while preserving coherence in system decisions;
- A behaviour selection mechanism expresses the global behaviour as composed of a repertoire of simple and independent motor patterns;
- Attention focusing, guided by behaviour intention, reduces preprocessing time of incoming perception data.

Each description of a stage is followed by its analysis according to memory requirement, processing complexity, and difficulty of modelling. Further discussions regarding robot programming and cognitive modelling points of view are also presented.

Finally, we describe an implementation on a mobile robot. The results demonstrate that the proposed framework is adequate for practical purposes.

Résumé

Les systèmes sensorimoteurs autonomes, placés dans des environnements dynamiques, doivent répondre continuellement à la question ultime : comment contrôler les commandes motrice à partir des entrées sensorielles ?

Répondre à cette question est un problème très complexe, principalement à cause de l'énorme quantité d'information qui doit être traitée, tout en respectant plusieurs restrictions : contraintes de temps réel, espace mémoire restreint, et capacité de traitement des données limitée. Un autre défi majeur consiste à traiter l'information incomplète et imprécise, habituellement présente dans des environnements dynamiques.

Cette thèse s'intéresse au problème posé par la commande des systèmes sensorimoteurs autonomes et propose un enchaînement d'hypothèses et de simplifications. Ces hypothèses et simplifications sont définies dans un cadre mathématique précis et strict appelé programmation bayésienne, une extension des réseaux bayésiens. L'enchaînement se présente en cinq paliers :

- les états internes résument la situation sensorimotrice et simplifient la modélisation, et, en conséquence, limitent la dégradation exponentielle de la performance à cause de la dimensionnalité ;
- l'hypothèse de Markov de premier ordre, l'hypothèse de stationnarité et les filtres bayésiens réduisent la dépendance au temps, sans négliger l'influence du passé ;
- l'indépendance partielle entre différents centres d'intérêt peut être exploitée pour réduire encore plus la dimensionnalité du problème tout en préservant la cohérence des décisions de système ;
- le mécanisme de choix de comportement exprime le comportement global comme une composition de modèles moteur simples et indépendants ;
- la focalisation de l'attention, guidée par intention de comportement, réduit le temps de pré-traitement des données relatives à la perception.

La description de chaque étape est suivie de son analyse selon les exigences de mémoire, de complexité de calcul, et de difficulté de modélisation. Nous présentons également des discussions approfondies concernant d'une part la programmation des robots et d'autre part les systèmes cognitifs.

Enfin, nous décrivons l'application de ce cadre de programmation à un robot mobile, dont les résultats démontrent qu'il est en adéquation avec des réalisations pratiques.

Resumo

Sistemas sensori-motor autônomos, situados em ambientes dinâmicos, devem continuamente responder à pergunta : como controlar os comandos motores conhecendo as entradas sensoriais ?

Responder à esta pergunta revela-se um problema extremamente complexo, principalmente devido ao enorme fluxo de informações a ser tratado enquanto diversas limitações devem ser respeitadas : restrições de tempo real, espaço de memória e poder de processamento limitados. Um desafio adicional é de tratar as informações incompletas e imprecisas, habitualmente presentes em ambientes dinâmicos.

Este trabalho de doutorado trata do problema de controlar sistemas sensori-motor autônomos e propõe uma sucessão de hipóteses e de simplificações acumulativas, que são definidas utilizando uma abordagem matemática precisa e estrita chamada Programação Bayesiana, uma extensão de redes bayesianas. Esta sucessão consiste de cinco etapas :

- Estados internos resumem a situação sensori-motor, simplificando a modelagem e reduzindo a degradação exponencial no desempenho causada pela grande quantidade de dimensões ;
- As hipóteses de Markov de primeira ordem, de estacionaridade e os filtros bayesianos reduzem a dependência no tempo, sem negligenciar a influência de informações passadas ;
- A independência parcial entre diferentes domínios de interesse pode ser explorada para reduzir ainda mais o número de dimensões do problema, preservando a coerência das decisões do sistema ;
- O mecanismo de seleção do comportamento expressa o comportamento global do sistema como uma composição de padrões motores simples e independentes ;
- A focalisação da atenção , guiada pela intenção do comportamento, reduz o tempo de tratamento dos dados provenientes da percepção .

A descrição de cada etapa é seguida de sua análise de acordo com a exigências de memória, complexidade de processamento e dificuldades involvidas na modelagem. Discussões relativas à programação de robôs e aos aspectos cognitivos relacionados também são apresentadas.

Uma aplicação prática em um robô móvel é descrita e seus resultados são considerados evidências de que a estrutura proposta mostra-se adequada para finalidades práticas.

Verdes mares bravios de minha terra natal, onde canta a jandaia nas frondes da carnaúba ;
Verdes mares, que brilhais como líquida esmeralda aos raios do sol nascente, perlongando as alvas praias ensombradas de coqueiros ;
Serenai, verdes mares, e alisai docemente a vaga impetuosa, para que o barco aventureiro manso resvale à flor das águas.

Vertes mers sauvages de mon pays natal, où chante le jandaia sous les frondaisons du palmier ;
Vertes mers, qui brillent comme émeraude liquide aux rayons du soleil levant, le long des plages blanches sous l'ombre des cocotiers ;
Soyez tranquilles, vertes mers, et lissez doucement la vague impétueuse, pour que le bateau aventureux glisse placide à fleur des flots.

José de Alencar, *Iracema*

A Viviane, com um sorriso.

Acknowledgements

"No man is an island", and this work would not be accomplished without the help of several persons. I would like to acknowledge their collaboration and support and there is no better way than to speak in their own language.

Je tiens à remercier Emmanuel Mazer pour avoir accepté de m'encadrer, pour sa motivation et sa patience. Pierre Bessière a investi beaucoup de son temps et de son énergie dans nos discussions : au-delà de ce qu'on attend d'un encadrant, sa bonne humeur et son amitié étaient présentes dans tous les moments.

I would like to thank Prof. Leslie P. Kaelbling for carefully reading and commenting the document as a member of my thesis committee. Jacques Droulez a accompagné mes travaux dès leurs premiers développements ; je lui suis reconnaissante pour son intérêt et son enthousiasme, et pour avoir accepté d'être rapporteur.

Meus sinceros agradecimentos ao Prof. Guilherme Carvalho pela leitura cuidadosa e pelas interessantes discussões . Jean Luc Schwartz m'a fait l'honneur de présider mon jury de thèse, et je tiens a lui remercier.

CAPES and Brazilian Ministry of Education financed me personally, and this work was also supported by the European Project BIBA. INRIA Rhône-Alpes provided physical structure and organisation. Without the help of SED, robot experimentations would not been achieved. Un grand merci à Soraya Arias pour son engagement personnel. J'adresse également mes remerciements à Veronique Roux, Anne Pasteur et Carole Bienvenu, pour leur aide diligente et inestimable, ainsi qu'à tout le personnel de l'INRIA Rhône-Alpes.

Un grand merci au personnel de l'équipe e-Motion pour l'ambiance d'échange, les discussions, l'amitié et la bonne humeur, spécialement dans les moments difficiles. Je remercie Francis pour sa patience, son oreille attentive et pour savoir dire le mot juste au bon moment. L'aide de Ronan dans la partie graphique du simulateur a été inestimable, et son indication de la suite pour violoncelle de Bach aussi. Albino a été le grand maître du robot BIBA et sans sa motivation, son engagement et ses capacités personnelles à résoudre les imprévus, la démo sur le robot n'aurait pas vu le jour.

Un merci spécial à mes chers amis de Grenoble, qui ont été ma famille pendant ces quatre ans. Kamel, Alicia et Jorge, Aracely et German, Azucena et Felipe, Angelica et Olivier, Mavi et Igor, Olivier Malrait, Kenneth, Iranete et Daniel, Anne Spanlanzani, parmi autant d'autres : vous êtes une partie importante de ma joie de vivre en France.

Gostaria de agradecer meus amigos que, mesmo fisicamente distantes, souberam es-

tar presentes durante minha estadia na França, especialmente Sérgio, Eneida, Alexandre, Carla, Solange, Mituhiro, Akemi e Annemarlen. Obrigada por me dar a sensação de ser alguém especial. Billy, Dolores, James and Paul were my regular company during writing : they helped me to stay awake.

Minha família sempre apoiou minhas idéias e decisões : meus pais, meu irmão e Marly sempre fizeram parte dos meus sonhos e da minha realidade ; Marieta, Helena e Assis me adotaram sem condições . Obrigada por acreditarem em mim, até quando eu mesma duvido.

Tetsu participou deste trabalho como revisor, conselheiro, amigo, psicólogo, família, apoio logístico e financeiro, e de muitas outras formas. Sem ele, dificilmente eu teria começado esta tese mas certamente não a teria acabado. Obrigada.

Contents

Notation **xxix**

Extended Abstract (in French) **xxxiii**

1 The Ultimate Question for Autonomous Sensory-Motor Systems **1**
1.1 A Naive Answer to the Ultimate Question 2
1.2 Requirements 4
1.2.1 Space Complexity 5
1.2.2 Time Complexity 5
1.2.3 Programming 5
1.3 Proposed Solution 5
1.4 Cognition 7
1.5 Contributions and Document Structure 8

2 Bayesian Framework for Dealing with Uncertainty **11**
2.1 Bayesian Approach for Reasoning and Expressing Uncertainty 12
2.1.1 Bayesian Robot Programming 12
2.2 Internal Variables 21
2.3 Conditional Independence of Variables 22
2.4 Decision Making 22
2.5 Summary and Open Questions 23

3 State of the World **25**
3.1 Bayesian Program 25
3.1.1 Description 26
3.1.2 Utilisation 27
3.1.3 Illustration Example 28
3.1.4 Additional Main Assumptions 30
3.2 Requirements Analysis 32
3.2.1 Space Complexity 33
3.2.2 Time Complexity 34
3.2.3 Programming 35
3.3 Discussion 35

3.3.1 Robotics . . . 35
3.3.2 Cognition . . . 36
3.4 Summary and Open Questions . . . 37

4 Bayes Filter **39**
4.1 Bayesian Program . . . 39
4.1.1 Description . . . 40
4.1.2 Utilisation . . . 41
4.1.3 Illustration Examples . . . 46
4.1.4 Additional Main Assumptions . . . 55
4.2 Requirements Analysis . . . 57
4.2.1 Space Complexity . . . 57
4.2.2 Time Complexity . . . 59
4.2.3 Programming . . . 61
4.3 Discussion . . . 61
4.3.1 Robotics . . . 61
4.3.2 Cognition . . . 64
4.4 Summary and Open Questions . . . 66

5 Exploiting Conditional Independence **69**
5.1 Bayesian Program . . . 69
5.1.1 Description . . . 70
5.1.2 Utilisation . . . 72
5.1.3 Illustrative Example . . . 74
5.1.4 Additional Main Assumptions . . . 77
5.2 Requirements Analysis . . . 78
5.2.1 Space Complexity . . . 78
5.2.2 Time Complexity . . . 80
5.2.3 Programming . . . 81
5.3 Discussion . . . 82
5.3.1 Robotics . . . 82
5.3.2 Cognition . . . 83
5.4 Summary and Open Questions . . . 85

6 Behaviour Coordination **89**
6.1 Bayesian Program . . . 89
6.1.1 Description . . . 90
6.1.2 Utilisation . . . 92
6.1.3 Illustration Example . . . 95
6.1.4 Additional Main Assumptions . . . 101
6.2 Requirements Analysis . . . 103
6.2.1 Space Complexity . . . 103
6.2.2 Time Complexity . . . 106

6.2.3 Programming . . . 108
6.3 Discussion . . . 109
6.3.1 Robotics . . . 109
6.3.2 Cognition . . . 116
6.4 Summary and Open Questions . . . 118

7 Attention Selection 121
7.1 Bayesian Program . . . 121
7.1.1 Description . . . 122
7.1.2 Utilisation . . . 123
7.1.3 Illustration Example . . . 127
7.1.4 Additional Main Assumptions . . . 133
7.2 Requirements Analysis . . . 135
7.2.1 Space Complexity . . . 135
7.2.2 Time Complexity . . . 139
7.2.3 Programming . . . 142
7.3 Discussion . . . 143
7.3.1 Robotics . . . 143
7.3.2 Cognition . . . 145
7.4 Summary and Open Questions . . . 147

8 Robot Experiments 149
8.1 Experimental Setup . . . 149
8.1.1 Behaviour Specification . . . 150
8.1.2 BIBA Robot Description . . . 150
8.1.3 Behaviour Implementation in BIBA robot . . . 152
8.2 Bayesian Programs . . . 154
8.2.1 Description . . . 155
8.2.2 Utilisation . . . 167
8.2.3 Safe Execution of Motor Commands . . . 170
8.3 Results . . . 173
8.3.1 Chase and Capture . . . 174
8.3.2 Obey with Obstacle Avoidance . . . 174
8.3.3 Reactivity to Predator . . . 174
8.3.4 Behaviour Coordination . . . 175
8.3.5 Result Analysis . . . 175

9 Conclusion and Perspectives 181
9.1 Contributions and Main Results . . . 181
9.2 Perspectives . . . 185
9.2.1 Hierarchical Bayesian Program . . . 187
9.3 Conclusion . . . 190

I Mathematical Details on Prediction-Estimation Loop 201

A Introduction 203
A.1 Proof Outline . . . 203
A.1.1 Recursive Calculation . . . 204
A.1.2 Equivalence between Elementary and Global Filters . . . 205
A.2 Tools and Rules . . . 206
A.2.1 General Inference Rule . . . 206
A.2.2 Normalisation . . . 206

B Elementary Filters 209
B.1 Elementary Filter . . . 209
B.1.1 Joint Distribution . . . 209
B.1.2 Prediction . . . 210
B.1.3 Estimation . . . 211
B.1.4 Motor . . . 212
B.1.5 Prediction using estimation . . . 213
B.2 Global Filter . . . 214
B.2.1 Joint Distribution . . . 214
B.2.2 Prediction . . . 215
B.2.3 Estimation . . . 216
B.2.4 Motor Commands . . . 217
B.2.5 Prediction using Estimation . . . 218
B.3 Relations Between Elementary and Global Filter . . . 219

C Behaviour Selection 225
C.1 Elementary Filter . . . 225
C.1.1 Joint Distribution . . . 225
C.1.2 Prediction . . . 226
C.1.3 Behavior Selection . . . 227
C.1.4 Estimation . . . 229
C.1.5 Motor Commands . . . 230
C.1.6 Prediction related to Estimation . . . 231
C.2 Global Filter . . . 232
C.2.1 Joint Distribution . . . 232
C.2.2 Prediction . . . 232
C.2.3 Behavior Selection . . . 234
C.2.4 Estimation . . . 236
C.2.5 Motor Commands . . . 237
C.2.6 Prediction using Estimation . . . 238
C.3 Relations Between Elementary and Global Filter . . . 239

D Attention Selection **247**
D.1 Elementary Filter . . . 247
D.1.1 Joint Distribution . . . 247
D.1.2 Prediction . . . 247
D.1.3 Attention Selection . . . 249
D.1.4 Behavior Selection . . . 251
D.1.5 Estimation . . . 252
D.1.6 Motor Command . . . 254
D.1.7 Prediction using estimation . . . 255
D.2 Global Filter . . . 256
D.2.1 Joint Distribution . . . 256
D.2.2 Prediction . . . 256
D.2.3 Attention Selection . . . 258
D.2.4 Behavior Selection . . . 260
D.2.5 Estimation . . . 262
D.2.6 Motor Command . . . 263
D.2.7 Prediction using Prediction . . . 264
D.3 Relations Between Elementary and Global Filter . . . 266

E Hierarchy of Behaviours **275**
E.1 Elementary Filter . . . 275
E.1.1 Joint Distribution . . . 275
E.1.2 Prediction . . . 275
E.1.3 Attention Selection . . . 282
E.1.4 Behavior Selection . . . 292
E.1.5 Estimation . . . 300
E.1.6 Motor Commands . . . 309
E.1.7 Prediction Loop . . . 312
E.1.8 Attention and Behaviour at level h, with $2 \leq h \leq L$ . . . 312
E.2 Global Filter . . . 315
E.2.1 Joint Distribution . . . 315
E.2.2 Prediction . . . 315
E.2.3 Attention Selection . . . 322
E.2.4 Behavior Selection . . . 332
E.2.5 Estimation . . . 339
E.2.6 Motor Commands . . . 348
E.2.7 Prediction Loop . . . 350
E.2.8 Attention and Behaviour at level h, with $2 \leq h \leq L$ . . . 351
E.3 Relations Between Elementary and Global Filter . . . 354
E.4 Requirements Analysis . . . 368
E.4.1 Space Complexity . . . 368
E.4.2 Time Complexity . . . 369

II Implementation Details **371**

F Simulation Examples **373**
F.1 Behaviour Selection Model . . . 373
F.2 Attention Selection Model . . . 374
F.2.1 Predator Filter . . . 374
F.2.2 Home Filter . . . 374

G Image Processing Algorithm **375**
G.1 Requirements . . . 375
G.2 Vision Processing . . . 375
G.2.1 Detection Algorithm . . . 375

H Robot Experiments **377**
H.1 Types Description . . . 377
H.1.1 Common Variables . . . 378
H.1.2 Programming Behaviour Model . . . 378
H.2 Prey Elementary Filter . . . 379
H.2.1 Relevant Variables . . . 379
H.2.2 Joint Distribution . . . 379
H.2.3 Parametrical Forms and Parameters Identification . . . 379
H.3 Trainer Elementary Filter . . . 382
H.3.1 Relevant Variables . . . 382
H.3.2 Joint Distribution . . . 382
H.3.3 Parametrical Forms and Parameters Identification . . . 383
H.4 Predator Elementary Filter . . . 385
H.4.1 Relevant Variables . . . 385
H.4.2 Joint Distribution . . . 385
H.4.3 Parametrical Forms and Parameters Identification . . . 386
H.5 Escape Route Elementary Filter . . . 387
H.5.1 Relevant Variables . . . 387
H.5.2 Joint Distribution . . . 388
H.5.3 Parametrical Forms and Parameters Identification . . . 388
H.6 Obstacle Avoidance . . . 390
H.7 Utilisation . . . 392

List of Figures

1.1 Naive solution for answering the ultimate question at time instant $t = 1$: one observation variable Z and one motor variable M. 3
1.2 Joint distribution of proposed framework. See text for explanation about color boxes. 6

2.1 Bayesian Program General Structure 16
2.2 Robot and farthest direction 17
2.3 Bayesian Program for following farthest direction 18
2.4 Farthest direction program. Top: curves for mean and standard deviation. Bottom: family of normal distributions for translation speed. 20
2.5 Farthest direction program. Top: curves for mean and standard deviation. Bottom: family of normal distributions for rotation speed when distance $Z_{dist} = 2$. 21

3.1 Top view of robot and position of ambient light sensors [Lebeltel, 1999]. . 28
3.2 Fusion of light sensors for light direction estimation. All curves show probability distributions over values of the variables Θ_l. The central curve is the result of fusion of the eight sensors. The sensor curves are disposed according to the positions of the sensors in the robot, and the value of each sensor measure is indicated on the top of the estimation. 31
3.3 Rotation speed for phototaxy behaviour. The curve shows the probability distribution for variable M_{rot} when the sensor measures are those shown in Figure 3.2 32

4.1 Diagram of Bayes filter utilisation. The dotted line indicates the transition from time instant $t - 1$ to time instant t. Expressions for prediction, estimation and motor questions can be found in the text. 46
4.2 Simulation of Example 1: robot (black) and predator (red). At left, variables describing the predator position and rotation speed signal convention are also shown. At right, the robot escaping from the predator. 47
4.3 States Prediction (left, in green) and Estimation (right, in blue) for time instant $t = 1$, simulation for Example 1, shown in right part of figure 4.2. . 50

4.4 Motor commands distribution for time instant t=1, simulation for Example 1, shown in right part of figure 4.2. 51
4.5 Illustration of simulation arena for Example 2, showing robot (black), predator (red), prey (green) and home (blue). 51
4.6 Sequence of snapshots for the simulation of Example 2. 55
4.7 Relationships between Bayesian approaches, from Diard [2003]. 62

5.1 Utilisation of elementary filters. 72
5.2 Sequence of snapshots for simulation of Example 2. On the left, simulation snapshots and, on the right, curves for rotation speed distributions, one for each filter answer and the resulting fusion. 87

6.1 Utilisation of Bayes filter with behaviour coordination. 92
6.2 Sequence of snapshots for simulation of the illustrative example. 100
6.3 Drawn values for behaviour variables for simulation of the illustrative example. 101
6.4 Behaviour selection (at left) and rotation speed (at right) at three different time instants. Behaviour selection curves show the answers of all elementary filters to the behaviour question and the resulting fusion. Rotation speed curves show the answers of all elementary filters to the motor question and the resulting fusion: the behaviour value employed in the motor question is the drawn value at each time instant. 102
6.5 States in the home filter for time instants $t = 30$ and $t = 31$. 116

7.1 Utilisation of Bayes filter with behaviour coordination and attention selection. The red arrow indicates that the behaviour probability distribution is predicted at the moment of the attention question. 124
7.2 Sequence of snapshots for simulation of Example 1. 131
7.3 Drawn values for the behaviour and attention variables for simulation of the illustrative example. 132
7.4 Curves for filter propositions and the resulting fusion for the behaviour variable at different time instants. 133
7.5 Predictions and estimations for predator state variables in illustrative example simulation. 134
7.6 Attention Selection in elementary filters and the resulting fusion in the illustrative example simulation, at time instant $t = 16$. 135
7.7 Attention Selection in elementary filters and the resulting fusion in the illustrative example simulation, at time instant $t = 39$. 136

8.1 The BIBA robot: at left, a side view; at right, a back view. 151
8.2 Robot BIBA and the computation structure employed in the experiments described in this chapter. 151
8.3 BIBA robot facing predator and trainer, at left. At right, the robot detects prey. 152

8.4 BIBA robot vision camera and laser capture actuator, with security cover. . 153
8.5 BIBA robot capturing prey. 154
8.6 Robot forward direction and objects direction. 155
8.7 Robot rotation speed zones and Robot translation speed. 156
8.8 Illustration of robot visible area and head position. 162
8.9 Utilisation of Bayes Filter for robot experiment. 169
8.10 Zones situated in front of the robot with an obstacle and detected distances 171
8.11 Rotation speed and translation speed terms in the zone bayesian program. 172
8.12 Chase behaviour. From top left to bottom right: (a) the robot wanders in environment, avoiding obstacles; (b) it detects a prey at its left side; (c) the robot advances towards prey; (d) the robot approaches nearest possible; (e) the prey is within reach; (f) the prey is captured with laser pointer . . . 175
8.13 Obey behaviour, with obstacle avoidance. From top left to bottom right: (a) the robot detects trainer; (b) the robot follows the trainer movements; (c) an obstacle comes between the robot and the trainer; (d), (e) the robot changes direction to avoid obstacle, but camera is always fixed on the trainer; (f) the robot advances towards the trainer. 176
8.14 Chase behaviour, with reactivity to predator appearance. From top left to bottom right: (a) the robot detects a prey; (b) the robot advances towards detected prey; (c) the robot detects a predator; (d), (e) escaping from detected predator; (f) hiding from detected predator (actually, blocked in a corner). 177
8.15 Behaviour coordination. From top left to bottom right: (a), (b) the robot wanders; (c) it detects a far predator; (c), (d) it keeps motionless; (e) approaching trainer detected; (f),(g),(h) the robot follows detected trainer; (i) even if a predator is very near, stays with the trainer. 178

9.1 Joint distribution of the proposed framework. The green square indicates the dynamic model; the blue square indicates the sensor model; the red square indicates the behaviour model; the orange square indicates the attention model; the brown square indicates the motor commands model; and the yellow square indicates initial conditions. 181
9.2 Utilisation of Bayes filter with behaviour coordination and attention selection. 182
9.3 Hierarchical model for defence behaviour applying the behaviour systems approach (see Timberlake and Silva [1995] for structure description). 187
9.4 Definition of one level in the hierarchy. 189
9.5 Utilisation of hierarchy levels in an elementary filter. 190

B.1 Bayes Network for Elementary Filter defined in joint distribution equation B.1. 210
B.2 Bayes Network for Global Filter in joint distribution in equation B.7. . . . 215

C.1 Bayes Network for Elementary Filter defined in joint distribution equation C.1. 226
C.2 Bayes Network for Global Filter defined in joint distribution equation C.7. 233

D.1 Bayes Network for Elementary Filter defined in joint distribution D.1. . . 248
D.2 Bayes Network for Global filter defined in joint distribution D.8. 257

E.1 Bayes Network for Elementary Filter defined in joint distribution E.1 for three levels (L=3). 276

List of Tables

1.1 Approximate numbers of probability values to be stored in relation to number of motor and observation variables, for $t = 8$ and $n = 16$: the expression is $2^{32\times(N_m+N_z)}$, where N_z is the number of observation variables and N_m is the number of motor variables. All variables are considered to have the same number of possible values, denoted n. 4

4.1 Approximate number of probability values to be stored as a function of the number of states, motor and observation variables. All variables are considered to have the same number of possible values, noted n. 57

6.1 Probability table for $P(B^j|B^{j-1}X^j_{dist}\pi_{pred})$, behaviour selection inverse model for predator filter. E: Escape behaviour; C: Chase behaviour; H: Homing behaviour; and M: Motionless behaviour. ++ means highly recommended, + means recommended, ∘ means not relevant, – means unwise and × means forbidden. An asterisk is employed when the situation is the same for all values of one variable. Red cell can be read as *If predator distance is three and robot was chasing it is forbidden to continue chasing*. 97

7.1 Probability table for $P(B^j|B^{j-1}Y^j_{dist}\pi_{prey})$, behaviour selection inverse model for prey filter. E: escape behaviour; C: chase behaviour; H: homing behaviour; and M: motionless behaviour. ++ means highly recommended, + means recommended, and ∘ means not relevant. 128

7.2 Probability table for $P(C^j|Y^j_{dist}B^t\pi_{prey})$, the attention selection inverse model for the prey filter. E: escape behaviour; C: chase behaviour; H: homing behaviour; and M: motionless behaviour. ++ means highly recommended, + means recommended, ∘ means not relevant, and × means forbidden. . . 129

7.3 Memory space and computational effort for illustrative examples in several chapters. 137

8.1 Probability table for $P(T^j_{pres}|T^{j-1}_{pres}\pi_{trai})$. 160

8.2 Probability table for $P(T^j_{pres}|T^{j-1}_{pres}[B^{j-1} = Obey]\pi_{trai})$. 160

8.3 Probability table for $P(ZT^j_{pres}|T^j_{pres}T^j_{dir}ZD^j_{pan}\pi_{trai})$, if $|T^j_{dir} - ZD^j_{pan}| \leq 2$. . . 161

8.4 Probability table for $P(B^j|B^{j-1}T^j_{pres}T^j_{dist}\pi_{trai})$, behaviour selection inverse model for the trainer filter. E: escape behaviour; C: chase behaviour; O: obey behaviour; and M: motionless behaviour. ++ means highly recommended, + means recommended, ∘ means not relevant, – means unwise and × stands for forbidden. An asterisk indicates that the situation is the same for all values of one variable. The red cell can be read as *If the robot was chasing and the trainer is visible it is recommended that the robot continues to chase*. 163
8.5 Probability table for $P(ZX^j_{pres}ZX^j_{timer}|X^j_{pres}X^j_{dir}ZD^j_{pan}\pi_{pred})$, if $|X^j_{dir}-ZD^j_{pan}| \leq 2$. 166
8.6 Probability table for $P(B^j|B^{j-1}X^j_{pres}X^j_{dist}\pi_{pred})$, behaviour selection inverse model for predator filter. E: Escape behaviour; C: chase behaviour; O: obey behaviour; and M: motionless behaviour. ++ means highly recommended, + means recommended, ∘ means unconcerned, – means unwise and × means forbidden. An asterisk indicates the situation is the same for all values of one variable. 167
8.7 Probability table for $P(B^j|E^j_{pres}\pi_{esc})$, behaviour selection inverse model for escape route filter. E: escape behaviour; C: chase behaviour; H: homing behaviour; and M: for motionless behaviour. ++ means highly recommended, + means recommended, ∘ means not relevant, – means unwise and × stands for forbidden. An asterisk indicates the situation is the same for all values of one variable. 168
8.8 Summary of strengths and shortcomings of proposed framework key points in the experiment described in this chapter. 180

F.1 Probability table for $P(B^j|B^{j-1}H^j_{dist}\pi_{home})$, behaviour selection direct model for home filter. E stands for Escape behaviour; C denotes Chase behaviour; H indicates Homing behaviour; and M signifies Motionless behaviour. ++ means highly recommended, + means recommended, ∘ means unconcerned, - means unwise and × stands for forbidden. An asterisk is employed when situation is the same for all values of one variable. 373
F.2 Probability table for $P(C^j|X^j_{dist}B^t\pi_{predator})$, attention selection direct model for predator filter. E stands for Escape behaviour; C denotes Chase behaviour; H indicates Homing behaviour; and M signifies Motionless behaviour. ++ means highly recommended, + means recommended, and ∘ means unconcerned. 374
F.3 Probability table for $P(C^j|H^j_{dist}B^t\pi_{home})$, attention selection direct model for home filter. E stands for Escape behaviour; C denotes Chase behaviour; H indicates Homing behaviour; and M signifies Motionless behaviour. ++ means highly recommended, + means recommended, and ∘ means unconcerned. 374

H.1 Probability table for $P(Y^j_{pres}|Y^{j-1}_{pres}\pi_{prey})$. 380

H.2 Probability table for $P(Y^j_{pres}|Y^{j-1}_{pres}[B^{j-1} = Chase]\pi_{prey})$. 380
H.3 Probability table for $P(ZY^j_{pres}|Y^j_{pres}Y^j_{dir}ZD^j_{pan}\pi_{prey})$, if $|Y^j_{dir} - ZD^j_{pan}| \leq 2$. . . 381
H.4 Probability table for $P(B^j|B^{j-1}Y^j_{pres}Y^j_{dist}\pi_{prey})$, behaviour selection direct model for prey filter. E stands for Escape behaviour; C denotes Chase behaviour; 0 indicates Obey behaviour; and M stands for Motionless behaviour. ++ means highly recommended, + means recommended, ∘ means unconcerned, - means unwise and × stands for forbidden. An asterisk is employed when situation is the same for all values of one variable. 381
H.5 Probability table for $P(T^j_{pres}|T^{j-1}_{pres}\pi_{trai})$. 383
H.6 Probability table for $P(T^j_{pres}|T^{j-1}_{pres}[B^{j-1} = Obey]\pi_{trai})$. 383
H.7 Probability table for $P(ZT^j_{pres}|T^j_{pres}T^j_{dir}ZD^j_{pan}\pi_{trai})$, if $|T^j_{dir} - ZD^j_{pan}| \leq 2$. . . 384
H.8 Probability table for $P(B^j|B^{j-1}T^j_{pres}T^j_{dist}\pi_{trai})$, behaviour selection direct model for trainer filter. E stands for Escape behaviour; C denotes Chase behaviour; 0 indicates Obey behaviour; and M stands for Motionless behaviour. ++ means highly recommended, + means recommended, ∘ means unconcerned, - means unwise and × stands for forbidden. An asterisk is employed when situation is the same for all values of one variable. 385
H.9 Probability table for $P(X^j_{pres}|X^{j-1}_{pres}\pi_{pred})$. 386
H.10 Probability table for $P(X^j_{pres}|X^{j-1}_{pres}[B^{j-1} = Escape]\pi_{pred})$. 386
H.11 Probability table for $P(ZX^j_{pres}ZX^j_{timer}|X^j_{pres}X^j_{dir}ZD^j_{pan}\pi_{pred})$, if $|X^j_{dir} - ZD^j_{pan}| \leq 2$. 387
H.12 Probability table for $P(B^j|B^{j-1}X^j_{pres}X^j_{dist}\pi_{pred})$, behaviour selection direct model for prey filter. E stands for Escape behaviour; C denotes Chase behaviour; H indicates Homing behaviour; and M stands for Motionless behaviour. ++ means highly recommended, + means recommended, ∘ means unconcerned, - means unwise and × stands for forbidden. An asterisk is employed when situation is the same for all values of one variable. 388
H.13 Probability table for $P(E^j_{pres}|E^{j-1}_{pres}\pi_{esc})$. 389
H.14 Probability table for $P(E^j_{pres}|E^{j-1}_{pres}[B^{j-1} = Escape]\pi_{esc})$. 389
H.15 Probability table for $P(B^j|E^j_{pres}\pi_{esc})$, behaviour selection direct model for escape route filter. E stands for Escape behaviour; C denotes Chase behaviour; H indicates Homing behaviour; and M stands for Motionless behaviour. ++ means highly recommended, + means recommended, ∘ means unconcerned, - means unwise and × stands for forbidden. An asterisk is employed when situation is the same for all values of one variable. 390

Notation

M: set of motor variables

Z: set of observation variables

S: set of state variables

B: set of behaviour variables

C: set of attention selection (focus of attention, selective perception) variables

N_m: number of variables for motor commands

N_z: number of observation variables

N_s: number of state variables

N_i: number of state subsets

N_{s_i}: number of state variables in subset i

N_{z_i}: number of observation variables in subset i

N_b: number of behaviour variables

N_c: number of attention selection (focus of attention, selective perception) variables

n: number of cases for discrete variables

$\mathcal{M}$: number of cells used in space complexity (memory requirements) analysis.

$\mathcal{C}$: number of mathematical operations to be executed. Used in time complexity analysis.

t: arbitrary discretised time value

j: index for discretised time

i: index for elementary filters

π: *a priori* knowledge for ultimate question

π_f: *a priori* knowledge of Bayes filter

π_s: *a priori* knowledge for sensor fusion

π_i: *a priori* knowledge for elementary filter i

π_{lf}: *a priori* knowledge for local filter presentation

π_{bs}: *a priori* knowledge for filter with behaviour selection

π_{as}: *a priori* knowledge for filter with attention selection (focus of attention, selective perception)

λ: set of coherence variables for motor command fusion

β: set of coherence variables for behaviour selection fusion

α: set of coherence variables for attention selection (focus of attention, selective perception) fusion

k: index for decision hierarchical level

L: number of decision levels in hierarchy

M^j: set of motor variables at time instant j

m^j: set of values for motor variables at time instant j

$m^{0:j}$ set of values for motor variables from time instant 0 to time instant j

${}_{[k]}C^j$: set of attention variables at level k for time instant j

${}_{[k]}c^j$: set of values for attention variables at level k for time instant j

${}_{[1:h]}C^j$: set resulting from the union of all attention variables at levels 1 to h, where $h > 1$ for time instant j:

$$ {}_{[k:h]}C^j = \bigcup_{k=1}^{h} {}_{[k]}C^j. $$

${}_{[1:L]}c^{0:j}$: set of values for attention variables at levels 1 to $h = L$, from time instant 0 to time instant j.

${}_{[1:L]}z^{0:j}$: set of values for observation variables of all elementary filters at levels 1 to L, from time instant 0 to time instant j:

$$ {}_{[1:L]}z^{0:j} = \bigcup_{k=1}^{L} \bigcup_{i=1}^{N_i} {}_{[k]}z_i^j . $$

ν_p^t: Set of known values of variables necessary for prediction at time instant t.

${}_{[h]}\nu_a^t$: Set of known values of variables necessary for attention selection at time instant t in level h.

${}_{[h]}\nu_b^t$: Set of known values of variables necessary for behaviour selection at time instant t in level h.

ν_m^t: Set of known values of variables necessary for the motor command at time instant t.

$\mathcal{G}(X), \mu, \sigma$: Normal curve over variable X, with mean μ and standard deviation σ.

Extended Abstract (in French)

Chapitre 1 : La Question Ultime

Quelles similitudes peuvent être trouvées entre un animal et un robot mobile autonome ? Tous deux doivent utiliser leurs capacités motrices en exploitant les informations acquises par leurs canaux perceptifs. Dans le cas d'un animal, les commande motrices sont principalement celles des muscles influant sur les articulations, et les informations sur l'environnement sont acquises par des organes de perception : les yeux, nez, oreilles, peau... Quant à un robot mobile, ses capacités motrices sont la plus part du temps des actionneurs électromécaniques ou pneumatiques, et l'information sur l'environnement provient des ses capteurs.

Dans ce travail de thèse, nous supposons qu'un système sensorimoteur autonome répond en permanence à la question suivante, dite question ultime : $P(M^t|z^{0:t} \wedge m^{0:t-1}\pi)$. Informellement, cette question peut être traduite par : *Qu'est-ce ce que je peux faire sachant ce que j'ai vu et ce que j'ai fait par le passé ?*

Mathématiquement, l'expression donne la distribution de probabilité sur les valeurs des variables de commande motrice M à l'instant t, sachant les valeurs des variables observées Z à partir de l'instant 0 jusqu'à l'instant t, ainsi que les valeurs de toutes les commandes motrices exercées entre l'instant 0 et l'instant $t-1$, et en prenant en compte les autres connaissances *a priori* du robot, de sa tâche et de son environnement(ces connaissances préalables sont assemblées et résumées par la variable π). Nous élaborons, dans les chapitres suivants de ce document, un cadre pour répondre à la question ultime. Ce cadre est appliqué à la programmation de comportements complexes pour un robot mobile. Nous défendons la thèse que l'utilisation de la focalisation de l'attention, et du choix de comportement, combinés au traitement de l'incertitude inhérente, contribuent à l'amélioration de la performance des systèmes de commande.

1.1 Réponse naïve à la question ultime

Cette section propose et discute une solution extrêmement simple pour répondre à la question ultime. Grâce à une analyse détaillée de cette première solution, notre objectif est de trouver des critères d'évaluation des différentes manières de répondre à la question ultime dans un système robotique.

La solution le plus simple consiste à créer un tableau de valeurs de probabilités indexé par toutes les différentes valeurs possibles des variables observées et des com-

mandes motrices. Chaque nouvelle variable ajoute une dimension et, au fur et à mesure que le temps passe, le nombre de dimensions du tableau augmentent encore plus vite. Chaque pas de temps additionnel ajoute autant de dimensions qu'il y a de variables sensorimotrices.

Cette solution très simple et basée sur une seule hypothèse (que toutes les variables, y compris le temps, sont discrètes), est-elle vraiment envisageable ?

Parmi plusieurs problèmes de faisabilité de cette solution, on peut citer l'énorme quantité de mémoire nécessaire pour garder ce tableau en mémoire. De plus il est extrêmement long de définir manuellement toutes les valeurs de probabilité requises. Ces points problématiques deviennent les principales motivations pour développer le cadre présenté dans ce document. L'analyse ici exécutée amène au choix des critères utilisés pour évaluer le cadre de programmation proposé et les différentes méthodes proposées.

1.2 Critères

1.2.1 Complexité spatiale

L'espace mémoire exigé pour garder toutes les données nécessaires d'une solution doit rester dans des limites raisonnables. Dans le cas d'un système robotique, il est possible utiliser quelques centaines de méga-octets de mémoire. Cependant, dans les chapitres suivants l'espace mémoire ne sera pas analysé numériquement mais plutôt par rapport à sa croissance asymptotique en fonction des paramètres de système (nombre de variables, par exemple). Ce critère s'appelle alors **Analyse de Complexité Spatiale**.

1.2.2 Complexité temporelle

Dans la solution naïve, un accès dans une table est suffisant pour retrouver une probabilité. Mais pour d'autres méthodes plus avancées, des calculs sont nécessaires. Le temps pris par ces calculs doit aussi être limité. Comme pour l'analyse de l'espace mémoire, le temps consommé dans les calculs sera analysé par sa variation asymptotique par rapport aux paramètres de taille du système. Ce critère s'appelle **Analyse de Complexité Temporelle**.

1.2.3 Programmation

Quel que soit le cadre de programmation employé, la programmation d'un comportement désiré implique certainement des prises de décisions et des ajustements de paramètres libres.

Ce critère vise à évaluer les difficultés de ces choix, les difficultés de modélisation d'un comportement spécifique dans le cadre proposé. Ceci implique d'évaluer quels sont les choix possibles de conception, ainsi que d'évaluer le nombre et la sémantique des paramètres libres du système.

La facilité de bénéficier des méthodes d'apprentissage est aussi un important critère d'évaluation. Quand un apprentissage est possible, l'évaluation de la méthode doit en tenir compte, ainsi que des possibles difficultés de l'expérimentation appropriée à cet

apprentissage.

1.3 Solution proposée

Les chapitres 4 à 7 décrivent les solutions successives à la question ultime, ordonnées par complexité croissante. Cette section présente aussi brièvement la solution proposée dans le chapitre 7, la dernière et la plus complexe des solutions. Elle utilise un filtre bayésien amélioré, incluant perception sélective, choix de comportement et sélection des commandes motrices.

1.4 Cognition

On suppose que la question ultime des systèmes sensorimoteurs est posée et résolue par les systèmes autonomes biologiques. Cette thèse présente des considérations mathématiques qui peuvent se révéler utiles comme éléments de réflexion à la compréhension de al façon dont les êtres vivants répond à la question ultime.

Chaque chapitre présente une section de discussion sur quelques possibles implications si les hypothèses proposées dans ce chapitre étaient vrai pour les systèmes cognitifs.

1.5 Contributions et plan de lecture

La contribution principale de cette thèse est la proposition d'un cadre pour programmer des robots autonomes, comportant une approche bayésienne pour le choix de comportement et la perception sélective. Cette méthode de programmation est définie dans un cadre mathématique strict (présenté au chapitre 2) et fondée sur une succession croissante d'hypothèses et de suppositions, d'où d'autres contributions additionnelles émergent.

Il est difficile d'établir la relation directe entre les observations des capteurs et les commandes motrices, sauf pour des comportements très simples. Afin de simplifier le processus d'écriture de cette relation, un niveau d'abstraction est présenté dans le chapitre 3 pour modéliser les aspects les plus pertinents de l'environnement en utilisant des variables d'état.

La question ultime prend en considération toutes les observations et les commandes moteur passées, ce qui signifie que l'espace mémoire et le temps de calcul augmentent sans bornes. Une solution est de limiter l'utilisation des informations sur le passé à une profondeur raisonnable. Le chapitre 4 applique l'hypothèse de Markov de première ordre et celle de la stationnarité pour maintenir l'espace mémoire et le temps de calcul constants dans le temps.

Une contribution secondaire dans ce chapitre est l'introduction de l'inférence des commandes motrices dans la définition du filtre bayésien, à l'intérieur même de la boucle récursive du calcul.

Bien que la mémoire et l'effort de calcul exigé soient constants dans temps, ils sont toujours trop grands et non limités par rapport à la taille de système. L'indépendance inhérente entre les variables d'état peut être exploitée pour réduire la complexité tem-

porelle. La division de l'espace d'état en sous-ensembles disjoints et l'utilisation de la fusion avec cohérence sont suggérées au chapitre 5 pour améliorer passage à l'échelle du système.

Quand le comportement spécifié pour le robot devient plus complexe, l'association directe des caractéristiques de l'environnement aux commandes motrices est lourd et inefficace. L'insertion d'un autre niveau d'abstraction permet d'associer des schémas de commandes motrices à des valeurs de comportement. Le chapitre 6 porte sur l'addition d'un mécanisme de coordination du comportement dans le filtre bayésien, qui constitue une autre contribution secondaire de ce travail.

Le flux des données des capteurs provenant d'un environnement riche implique l'utilisation de méthodes de traitement de ces données pour en extraire les informations désirées. En conséquence, un temps est nécessaire pour traiter toutes ces données. Le chapitre 7 propose un mécanisme de choix d'attention pour réduire ce temps de traitement de données. La proposition d'un mécanisme bayésien de perception sélective associé au choix de comportement est une contribution importante de ce chapitre.

Des preuves de l'applicabilité des concepts proposés ci-dessus pour contrôler un robot sont données au chapitre 8. La méthode appliquée pour programmer cette expérimentation est une contribution additionnelle.

Le chapitre 9 récapitule les contributions, les résultats principaux et les perspectives.

Tous les chapitres proposant une solution sont organisés de la même façon. Au début, la proposition du chapitre est présentée sous la forme d'un programme bayésien, dont la présentation est elle-même composée d'une description, d'une utilisation, et d'un exemple illustratif. Les hypothèses principales de la proposition sont alors formellement présentées.

Puis la proposition du chapitre est analysée : les complexités en temps et en espace sont détaillées, suivies de l'analyse des difficultés de programmation.

On propose ensuite une discussion générale concernant les domaines de la robotique et des systèmes cognitives. Cette discussion porte sur d'autres travaux scientifiques et leurs différences par rapport à la proposition du chapitre (hypothèses, contraintes, portée et répercussions).

Le chapitre finit avec un bref sommaire et une évocation de questions à développer.

D'autres contributions secondaires de ce travail sont l'analyse des complexités temporelle et spatiale ainsi que l'analyse des aspects cognitifs de la description et de l'utilisation des filtres bayésiens proposés.

Chapitre 2 : Approche Bayésienne

Afin de répondre efficacement à la question ultime, nous nous servons de concepts et de méthodes pour exprimer nos hypothèses dans un cadre de programmation. Ce chapitre compile les outils théoriques et pratiques utilisés dans les chapitres suivants.

2.1 Approche bayésienne pour exprimer et raisonner avec incertitude

Plusieurs méthodes pour traiter les diverses sources d'incertitudes ont été propo-

sées, dont l'approche probabiliste, qui permet de représenter l'incertitude. Le théorème de Bayes est l'outil correspondant pour raisonner en présence d'incertitude.

2.1.1 Programmation bayésienne des robots

La programmation bayésienne est un cadre d'unification pour décrire formellement les modèles probabilistes. L'objet principal de la programmation bayésienne est le programme bayésien. Il se compose d'une *description* de connaissances préalables, de données expérimentales, et d'une phase d'*utilisation*.

Un exemple est proposé comme illustration où l'on désire qu'un robot suit la direction la plus lointaine détectée avec un capteur laser SICK.

2.2 Variables d'état

Dans ce travail les variables d'état rempliront différentes fonctions : décrire les caractéristiques de l'environnement, raisonner, déduire des décisions internes, exprimer des quantités inhérentes à chaque comportement, etc.

2.3 Indépendance conditionnelle des variables

Dans un modèle quelconque, l'hypothèse la moins restrictive est que, *a priori*, toutes les variables sont dépendantes Mais, en pratique, particulièrement avec un nombre très élevé de variables, considérer toutes les dépendances possibles entre les variables implique une complexité ingérable. Dans ces situations, l'indépendance conditionnelle est utile pour réduire la complexité de problème. Par exemple, deux variables peuvent être considérées indépendantes sachant la valeur d'une troisième.

2.4 Prise de décision

Dans un système comme celui proposé ici pour commander un robot, des nombreuses variables sont employées, et si, pour certains d'entre elles la propagation d'incertitude est essentielle, pour d'autres elle n'est pas importante et peut même être gênante. Pour ces dernières, une décision peut être prise sur leurs valeurs, arrêtant le processus de propagation d'incertitude : ces variables n'ont maintenant plus de distributions de probabilité mais des valeurs attribuées fixes.

Chapitre 3 : État du Monde

Ce chapitre traite de l'addition des variables appelées *états*, généralement utilisées en robotique. Ces variables concernent les caractéristiques de l'environnement les plus pertinentes à l'exécution du comportement désiré, et elles sont également liées aux caractéristiques dynamiques du robot (comme la position, la direction, et la vitesse, par exemple).

L'utilisation des variables d'état simplifie l'utilisation des signaux des capteurs et la définition des commandes moteur, désaccouplant les variables d'observation et les

variables de commandes moteur.

3.1 Programme bayésien

Ce programme bayésien montre comment ajouter des variables d'état pour désaccoupler les observations de capteurs et des commandes de moteur. En outre, le schéma de fusion bayésienne est présenté comme une façon d'améliorer le calcul des distributions de probabilité de variables d'état en combinant des évidences de plusieurs observations des capteurs.

3.1.1 Description

En plus de l'ensemble de variables d'observation Z et de l'ensemble de variables motrices M, ce programme inclut des variables internes d'état qui constituent l'ensemble de variables pertinentes et reliées au robot, à la tâche indiquée et aux caractéristiques les plus significatives dans l'environnement du robot. Toutes ces variables sont considérées à l'instant j. De plus, Z^j (l'ensemble des variables d'observation à l'instant j) est défini comme la composition de N_z variables d'observation :

$$Z^j = \{{}_1Z^j, {}_2Z^j, {}_{N_z}Z^j \ldots\}, \forall j. \tag{1}$$

La distribution conjointe sur les variables numérotées ci-dessus est définie comme :

$$\begin{aligned} & P({S^j}_1Z^j{}_2Z^j \ldots {}_{N_z}Z^j M^j | \pi_s) \\ = & \left| \begin{array}{l} P(S^j|\pi_s) \\ \prod_{k=1}^{N_z} [P({}_kZ^j|S^j\pi_s)] \\ \times P(M^j|S^j\pi_s). \end{array} \right. \end{aligned} \tag{2}$$

3.1.2 Utilisation

Ce programme peut être employé pour estimer les variables d'état sachant les mesures des capteurs. Une autre utilisation du programme présenté ci-dessus est de trouver la distribution de probabilité sur les variables commandes motrices sachant les mesures des capteurs.

3.1.3 Exemple

Un exemple est présenté pour illustrer le programme bayésien décrit ci-dessus. Nous considérons un robot doté de deux roues et avec huit capteurs de lumières ambiantes. On désire que ce robot exécute un comportement de phototaxie. Cet exemple montre comment la distribution conjointe est construite, ainsi que les questions posées à cette distribution conjointe et des résultats de simulation.

3.1.4 Hypothèses supplémentaires

- Les variables d'état désaccouplent les variables d'observation des capteurs et les variables de commandes motrices ;
- les observations résultent du traitement des mesures des capteurs, et peuvent être exprimées en fonction des variables d'état ;
- Les variables d'observations des capteurs sont supposées indépendantes sachant les variables internes d'état.

3.2 Analyse de critères

Cette analyse vise à souligner les avantages apportés par l'emploi des variables d'état.

3.2.1 Complexité spatiale

Maintenir ce programme en mémoire signifie stocker toutes les distributions de probabilités présentes dans l'équation 3, ainsi que les réponses aux questions.

L'espace nécessaire pour garder ce programme bayésien en mémoire augmente selon la taille du modèle capteur et du modèle moteur. Cet espace augmente de manière linéaire avec le nombre de variables d'observation et de manière exponentielle avec le nombre de variables d'état et des variables motrices.

3.2.2 Complexité temporelle

Deux questions sont posées à cette distribution commune et, pour évaluer la complexité temporelle de ce programme bayésien, il est nécessaire de savoir combien d'opérations sont nécessaires pour calculer chaque réponse à ces questions.

Le temps nécessaire augmente linéairement avec le nombre de variables d'observation, et exponentiellement avec le nombre de variables d'état et de variables motrices.

3.2.3 Programmation

Comme les variables d'état décrivent la plupart des caractéristiques pertinentes à la tâche désirée du robot, on s'attend à ce que ces variables soient suffisantes pour définir des commandes motrices. Ainsi, il est plus facile d'associer les commandes motrices avec les variables d'état qu'avec ces variables d'observation.

Les observations se rapportent à plusieurs aspects du robot et de l'environnement dont certains ne sont même pas pertinents pour la tâche indiquée de robot. Les modèles capteur nous permettent de se concentrer sur les aspects significatifs des et leur relation avec les variables d'état.

3.3 Discussion

3.3.1 Robotique

Les variables d'état et la fusion de capteur sont employées régulièrement en robotique mobile pour plusieurs raisons. Les variables d'état peuvent désaccoupler des sys-

tèmes de capteur et de moteur, permettant de ce fait un raisonnement plus sophistiqué et des comportements complexes conséquents.

Pour dresser la carte des environnements d'un robot, par exemple, les variables d'état décrivent les différents lieux. L'avantage principal de la fusion capteur est d'augmenter la qualité de l'estimation des états. La redondance des observations améliore la précision de l'estimation et assure la robustesse aux défaillances.

3.3.2 Cognition

Les états internes utilisés par un observateur pour modéliser un animal observé sont probablement complètement diffèrents des états internes utilisés pour l'animal lui même.

Nous suggérons que l'observateur et l'animal observé ont, toutes les deux, des états internes parce que l'utilisation de ceux-ci permet une réduction substantielle des besoins de mémoire et de complexité de traitement. Mais, comme les deux ont des entrées sensorielles diverses, ils ont probablement des états internes différenciés.

Chapitre 4 : Filtre Bayésien

Ce chapitre propose appliquer une structure de filtre bayésien afin de désaccoupler partiellement la dépendance entre les variables à différents instants. On ne souhaite pas oublier toute l'information du passé, mais limiter la dépendance du temps à une profondeur raisonnable. Cette limite est imposée ici en appliquant l'hypothèse de Markov de premier ordre.

La dépendance avec le temps est également réduite en assumant que les modèles sont stationnaires dans temps. Des informations sur le passé sont résumées et des considérations instantanées du pas de temps présent sont utilisées pour estimer les valeurs actuelles des variables. Cette estimation est réalisée par un calcul récursif qui réduit le temps de calcul.

Le filtre bayésien proposé dans ce chapitre inclut les commandes de moteur et son modèle dans le calcul récursif du filtre. Cela permet au filtre de fournir une réponse à la question ultime, et constitue ainsi une première proposition de solution.

4.1 Programme bayésien

Nous visons ici définir la dépendance entre les variables à un moment donné, et maintenons ces dépendances stationnaires : elles ne changent pas avec le temps. Afin de maintenir la relation avec le passé, les variables considérées dans chaque terme de la distribution conjointe incluent les variables actuelles et les variables de temps passé. La dépendance entre les variables utilisées dans le chapitre précédent est employée ici avec quelques différences : les modèles capteur et moteur sont encore employés, mais un modèle dynamique concernant les variables d'état est ajouté. Si nécessaire, les variables de temps passé sont incluses dans les dépendances des modèles. L'utilisation est également très similaire, avec des questions d'estimation des états et de commandes

motrices. Afin de réduire le temps de calcul, une boucle de calcul récursif est employée pour répondre à ces questions.

4.1.1 Description

Comme dans le chapitre précédent, les variables impliquées dans ce programme sont l'ensemble des variables d'observation Z, l'ensemble des variables M de commande de moteur et l'ensemble de variables d'état S. Mais, contrariament au chapitre précédent, tous les instants, de 0 à un instant arbitraire t, sont considérés. Ainsi l'ensemble total de variables pertinentes est $M^{0:t}$, $Z^{0:t}$ et $S^{0:t}$.

La distribution conjointe du filtre bayésien fait intervenir toutes les variables pertinentes :

$$\begin{aligned} & P(M^{0:t}S^{0:t}Z^{0:t}|\pi_f) \\ & = \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^jM^{j-1}\pi_f) \end{array} \right] \\ \times P(M^0S^0Z^0|\pi_f). \end{array} \right. \end{aligned} \tag{3}$$

On remarque que l'hypothèse de Markov de premier ordre est présente dans le modèle dynamique et dans le modèle moteur : la dépendance de temps a une profondeur limité à un instant. Le produit des termes dans la distribution conjointe est une conséquence de l'hypothèse de stationnarité : les modèles ne changent pas avec le temps. Ils sont définis pour différentes variables à chaque pas de temps, mais la relation entre ces variables demeure le même pour tous les pas de temps.

4.1.2 Utilisation

Le calcul récursif dans le filtre bayésien ci-dessus consiste à poser trois questions, séquentiellement, à la distribution conjointe du filtre. La première question s'appelle *prédiction*, et c'est la partie de la boucle de calcul où les informations sur le passé sont mises à jour en utilisant le modèle dynamique.

La deuxième question s'appelle *estimation*, et elle vise à mettre à jour le résultat de la question de prédiction avec l'évidence donnée par les observations.

Une fois que l'estimation des variables d'état est calculée, il est possible d'appliquer cette estimation pour décider quelles commandes de moteur sont les plus appropriées pour la situation actuelle : c'est la question de *commande motrice*.

Le calcul de la prédiction utilise l'estimation calculée au pas de temps précédent, l'estimation actuelle est calculée en utilisant la prédiction, et la question de commande motrice utilise l'estimation pour être calculée. La recursivité du calcul est démontrée dans cette section.

4.1.3 Exemples

Deux exemples sont présentés dans cette chapitre : le premier est une version réduite du deuxième.

Le premier exemple montre un filtre bayesien et son utilisation : le choix des variables, la distribution conjointe, la façon dont les formes paramétriques sont choisis, et les questions posées. Des résultats de la simulation sont aussi montrés pour illustrer le fonctionnement du filtre.

Le deuxième exemple est plus complexe et montre comment un comportement plus sophistiqué peut être programmé.

4.1.4 Hypothèses supplémentaires

- Les variables d'état du pas de temps actuel dépendent des états du pas de temps précédent aussi bien que des variables moteur du pas de temps précédent.
- L'hypothèse de Markov de première ordre impose une limite pour la profondeur de dépendance dans le temps à un pas de temps.
- Les modèles dynamiques, capteur et moteur sont stationnaires dans le temps.

4.2 Analyse de critères

4.2.1 Complexité spatiale

Pour comparer avec la solution proposé au chapitre 1 pour répondre à la question finale, un tableau présente l'espace mémoire nécessaire pour stocker le filtre bayésien proposé dans ce chapitre. Une réduction substantielle de place mémoire est visible.

Le problème principal de l'utilisation du filtre Bayésien est la dépendance exponentielle au nombre de variables d'état. Dans les applications en robotique, le nombre de variables d'état est directement lié à la difficulté du comportement spécifié et la sophistication de l'environnement : le nombre de variables d'état augmente avec la complexité de la tâche et avec la richesse de l'environnement.

4.2.2 Complexité temporelle

Le temps de calcul pour l'utilisation du filtre de Bayes est la somme des opérations nécessaires pour résoudre les trois questions à chaque pas de temps : prédiction, estimation et commande motrice.

Comme pour la complexité spatiale, nous trouvons encore une dépendance exponentielle à l'égard du nombre de variables d'état.

4.2.3 Programmation

Il est nécessaire de programmer trois termes de la distribution conjointe : modèle dynamique, modèle capteur et commande moteur. Ces termes ont une signification précise directement liée au robot, à l'environnement et à la tâche à programmer. Pour cette raison, même si une table est utilisée au lieu des formes paramétriques, chaque valeur de probabilité peut être associée à une situation réelle, et peut donc être choisie plus facilement.

4.3 Discussion

4.3.1 Robotique

Le programme bayésien décrit dans ce chapitre n'est pas classifié comme une spécialisation de la localisation Markovienne à cause de la présence du modèle moteur dans la distribution conjointe. Nous classifions notre proposition comme un type de filtre Bayesien.

Les variables motrices, non incluses dans la structure générale d'un filtre bayésien basique, sont utilisées dans notre proposition, et elles sont très similaires aux variables d'état.

De même que les variables d'état, les variables motrices ont une dynamique associée, donnée par le modèle moteur. Dans l'expression de la solution de la question moteur, on voit intervenir l'estimation des variables d'états. De ce fait, les variables motrices sont estimées, plutôt que predictes.

En fait, la principale différence entre les variables d'état et les variables motrices réside dans le fait que ces dernières ont des valeurs attribuées. À la fin de chaque itération, une décision est prise au sujet des valeurs des commandes motrices et, à la prochaine itération, les variables moteur ont des valeurs attribuées, alors que les variables d'état ne sont jamais connues avec précision, mais seulement estimées.

4.3.2 Cognition

Si on assume que le cerveau utilise un filtre bayésien, cela implique le cerveau utilise des distributions de probabilités. Cette hypothèse amène à la révision des plusieurs concepts dans l'étude de comment l'information est codée, processée et utilisé par le cerveau. Le besoin de prendre de décisions, l'horizon de temps et la relation entre modèles stationnaires et l'apprentissage sont aussi discutés.

Chapitre 5 : Indépendance Conditionnelle

Ce chapitre vise à réduire la complexité en exploitant l'indépendance conditionnelle entre les variables d'état et les variables d'observation. Cette indépendance existe en raison de l'existence de sous-ensembles disjoints dans les espaces d'état et d'observation. Les variables motrices, cependant, sont liées à toutes les autres variables, et aucune hypothèse d'indépendance ne peut être appliquée. Nous traitons cette situation en utilisant la fusion avec cohérence.

5.1 Programme bayésien

Dans ce chapitre, l'indépendance conditionnelle dans les espaces d'état et d'observation est exploitée pour réduire les sommes imbriquées. L'espace d'état est préservé, et des additions sont encore exécutées, mais elles peuvent être calculées indépendamment.

Tandis que l'indépendance conditionnelle peut exister dans les espaces d'état et d'observation, elle ne peut pas être appliquée aux variables de commande motrice. Un

modèle peut être écrit en employant des variables de cohérence, sous la forme d'une fusion avec cohérence, qui n'ajoute pas de nouvelles connaissances ni de nouvelles hypothèses. Cette méthode est utilisée dans ce chapitre pour définir le modèle moteur dans le programme bayésien proposé.

5.1.1 Description

Un nouveau type de variable est utilisé pour indiquer si l'ensemble de variables pertinentes dans un modèle ont des valeurs cohérentes entre elles. Dans ce chapitre, λ^t est un ensemble de variables binaires de cohérence pour le modèle de commande motrice.

La distribution conjointe du filtre bayésien sur toutes les variables pertinentes est :

$$\begin{aligned} &P(M^{0:t}S^{0:t}Z^{0:t}\lambda^{0:t}|\pi_{lf}) \\ &= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^j|S_i^{j-1}M^{j-1}\pi_i)\right] \\ \prod_{i=1}^{N_i} \left[P(Z_i^j|S_i^j\pi_i)\right] \\ \times P(M^j|\pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i)\right] \end{array} \right] \\ \times P(M^0S^0Z^0\lambda^0|\pi_{lf}). \end{array} \right. \end{aligned} \tag{4}$$

Le modèle de commande de moteur est maintenant défini en utilisant la fusion avec cohérence, et il implique que les variables λ de cohérence sont aussi assemblées en sous-ensembles mutuellement exclusifs.

La distribution conjointe ci-dessus définit un filtre appelé *filtre global*, et elle est définie sur toutes les variables pertinentes. Après la division de l'espace des variables d'état et d'observation en des sous-ensembles mutuellement exclusifs, il est possible de dériver l'existence des *filtres élémentaires*, un pour chaque sous-ensemble de variables :

$$\begin{aligned} &P(M^{0:t}S_i^{0:t}Z_i^{0:t}\lambda_i^{0:t}|\pi_i) \\ &= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\ \times P(Z_i^j|S_i^j\pi_i) \\ \times P(M^j|\pi_i) \times P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i) \end{array} \right] \\ \times P(M^0S_i^0Z_i^0\lambda_i^0|\pi_i). \end{array} \right. \end{aligned} \tag{5}$$

La troisième ligne montre le modèle moteur défini sur le sous-ensemble S_i de variables d'état et les variables motrices M. Elle est définie ici en utilisant la fusion avec cohérence.

5.1.2 Utilisation

Au lieu de poser des questions à la distribution conjointe du filtre global, nous souhaitons maintenant utiliser les filtres élémentaires autant que possible. L'utilisation consiste donc à poser les questions de prédiction, estimation et commande motrice aux distributions conjointes des filtres élémentaires. Pour avoir la distribution de probabilité

globale des variables de commande motrice, la fusion *a posteriori* des réponses de tous les filtres élémentaires doit être exécutée.

5.1.3 Exemple

Le deuxiéme exemple présenté dans le chapitre précèdent est repris ici et adapté pour l'utilisation des filtres élémentaires. Les résultats de simulation illustrent l'utilisation de fusion basée en cohérence pour la programmation des modèles de commande motrice.

5.1.4 Hypothèses supplémentaires

- Des variables d'état et d'observation peuvent être assemblées en des sous-ensembles mutuellement exclusifs.
- Il est plus pratique écrire un modèle direct pour des commandes moteur qu'un modèle inverse.

5.2 Analyse de critères

5.2.1 Complexité spatiale

Seuls les termes de la distribution conjointe des filtres élémentaires doivent être maintenues dans la mémoire, ainsi que les réponses aux questions posées à chaque filtre élémentaire.

La complexité spatiale augmente maintenant exponentiellement avec le nombre de variables d'état dans le plus grand sous-ensemble et linéairement avec le nombre de filtres élémentaires.

5.2.2 Complexité temporelle

De manière similaire à la complexité spatiale, la taille du plus grand sous-ensemble de variables d'état détermine la façon dont le temps de calcul augmente avec la taille de système.

5.2.3 Programmation

Comparé au filtre bayésien proposé dans le chapitre précédent, la programmation devient plus facile à l'intérieur de chaque filtre élémentaire. La condition de l'indépendance entre les variables d'état et d'observation est suffisante pour assurer que les modèles dynamique et capteur dans chaque filtre élémentaire puissent être isolés des autres filtres élémentaires.

La difficulté principale pour programmer cette version est le choix des variables qui remplissent les conditions des sous-ensembles mutuellement indépendants par rapport aux modèles dynamique et capteur.

5.3 Discussion

5.3.1 Robotique

La réduction de complexité temporelle des filtres bayésiens est le sujet de plusieurs projets de recherche, et quelques approches sont brièvement présentées et mises en relation avec notre proposition.

5.3.2 Cognition

Les caractéristiques de l'environment pertinentes au comportement peuvent être considérées séparément comme des centre d'intérêt. Cette séparation amène à des simplifications importantes et conséquente réduction de complexité.

Chapitre 6 : Coordination des comportements

Des commandes motrices peuvent être regroupées en motifs simples et indépendants : suivre un objet, aller tout droit, et ainsi de suite. Chaque motif est appelé ici un comportement de base, et dans ce chapitre nous proposons l'addition d'un mécanisme de choix de comportement au filtre bayésien du chapitre précédent. Le choix du comportement le plus adéquat est fait en utilisant la fusion des propositions des filtres élémentaires, qui sont le résultat d'une stratégie locale de choix de comportement de chaque filtre élémentaire selon son propre sous-ensemble de variables.

6.1 Programme bayésien

Pour programmer un robot de telle manière qu'il exécute un comportement préalablement indiqué, les combinaisons des motifs simples de commande motrice sont souvent suffisantes pour créer des comportements résultantes plus complexes.

Ce chapitre présente des commandes motrices regroupées en motifs, et chaque filtre élémentaire évalue chaque motif moteur possible afin d'estimer son adéquation avec son propre sous-ensemble de variables d'état et d'observation. La fusion des propositions de chaque filtre élémentaire permet de décider quel motif de commande motrice est plus convenable et permet son exécution cohérente avec l'aide des modèles moteur de tous les filtres élémentaires.

6.1.1 Description

Pour coordonner l'exécution des motifs moteur, nous proposons ajouter un ensemble de variables de comportement, notées B^t, communes à tous les filtres élémentaires. La sémantique de ces variables est liée aux motifs trouvés dans les commandes motrices : chaque variable correspond à un groupe de motifs moteur, une valeur pour chaque motif.

Ces variables de comportement sont communes à tous les filtres élémentaires, et la fusion avec cohérence est appliquée pour décrire le modèle de choix de comportement : β^t représente l'ensemble de variables de cohérence liées aux variables de comportement

B^t. Le nombre de variables de cohérence dépend du nombre de variables de comportement et du nombre de filtres élémentaires : une variable β est nécessaire pour chaque variable de comportement B dans chaque filtre élémentaire.

La distribution conjointe pour un filtre élémentaire général est montrée ci-dessous.

$$\begin{aligned} & P(M^{0:t} S_i^{0:t} Z_i^{0:t} B^{0:t} \lambda_i^{0:t} \beta_i^{0:t} \pi_i) \\ & = \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i). \end{array} \right. \end{aligned} \tag{6}$$

La troisième ligne montre le modèle de comportement, dans la forme de fusion avec cohérence $P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i)$. Le comportement choisi au pas de temps actuel dépend des variables d'état et du comportement choisi au pas de temps précédent.

La relation entre le comportement actuel et le comportement précédent vise à assurer la continuité (aussi appelée persistance) dans l'exécution d'un comportement. La dépendance à l'égard de l'état actuel favorise la réactivité aux changements dans l'environnement.

Il est intéressant de souligner le rôle de la variable de comportement dans le modèle moteur. Les variables de comportement exécutent la sélection du motif moteur : chaque valeur des variables de comportement correspond à un motif moteur défini par un modèle moteur spécifique.

6.1.2 Utilisation

L'utilisation de ce filtre consiste à poser les questions de prédiction, comportement, estimation et commandes motrices aux distributions conjointes des filtres élémentaires. Après avoir posé la question de comportement à tous les filtres élémentaires, la fusion des réponses des filtres est réalisée et à partir du résultat de la fusion, la valeur actuelle du comportement choisi est décidée. Au moment où la question de commande motrice est posée à chaque filtre, la valeur du comportement actuel est déjà connue et pour avoir la distribution de probabilité globale des variables de commande motrice, la fusion *a posteriori* des réponses de tous les filtres élémentaires doit être exécutée.

6.1.3 Exemple

L'exemple présenté dans les chapitres précédents est repris ici et on y ajoute un modèle de choix de comportement à partir de l'identification des motifs moteurs.

6.1.4 Hypothèses supplémentaires

- Des commandes motrices peuvent être assemblées dans les motifs bien identifiés, appelés comportements de base.

- La stratégie de choix de comportement est individuelle pour chaque filtre élémentaire. Des propositions de chaque filtre sont combinées pour un choix global.
- Le choix du comportement actuel est influencé par des caractéristiques de robot et d'environnement, et par le comportement spécifié pour le robot.
- Le choix de comportement dépend du comportement au pas de temps précédent.
- Un seul comportement est actif à la fois.

6.2 Analyse de critères

6.2.1 Complexité spatiale

La quantité de mémoire augmente avec le plus grand sous-ensemble de variables d'état et d'observation, avec le nombre de filtres élémentaires, et avec le nombre de variables de commande motrice et de comportement. Il est possible conclure aussi que la complexité de l'espace ne change pas avec l'ajout des variables de comportement pour la plupart des applications en robotique.

6.2.2 Complexité temporelle

Un nouveau terme dans la complexité temporelle apparaît, lié aux questions de comportement, et qui dépend du plus grand sous-ensemble de variables d'observation, des variables de comportement et du plus grand sous-ensemble de variables d'état.

Pour la plupart des applications en robotique, pourtant, il est possible conclure que l'addition des variables de comportement ne change pas l'expression de complexité du chapitre précedent.

6.2.3 Programmation

Le modèle de choix de comportement met alors en application la stratégie de choix dans chaque filtre élémentaire, sur la base de ses propres variables d'état, et par conséquent sans disposer des informations concernant les variables d'état dans d'autres filtres.

Il est plus facile de le programmer parce que des connaissances limitées sont utilisées dans chaque filtre, et un plus petit nombre de variables doit être pris en compte.

La programmation proscriptive met l'accent sur la nécessité d'interdire des comportements prohibés plutôt que de recommander des valeurs désirées. Elle est fortement désirable dans un schéma de fusion des capteurs, et dès que la fusion avec cohérence donne un résultat similaire, elle est également utile pour des commandes motrices et des modèles de choix de comportement.

Des niveaux de probabilité sont choisis pour classifier des comportements pour une situation indiquée : fortement recommandée, recommandée, indifférente, imprudente, strictement interdite. Chaque situation est définie en associant une valeur de probabilité pour chaque comportement possible en fonction de la valeur du comportement choisi au pas de temps précédent et des valeurs possibles des variables d'état actuelles. Au lieu d'avoir un très grand nombre de valeurs de probabilité à ajuster (approximativement égal à la quantité de mémoire nécessaire pour stocker le modèle de comportement

dans la mémoire), seulement cinq paramètres doivent être accordés : les valeurs de probabilité utilisées pour chaque situation décrite ci-dessus.

6.3 Discussion

6.3.1 Robotique

Nous discutons ici deux points principaux concernant des mécanismes de choix d'action : la granulité de la définition d'action, et les critères habituels pour un mécanisme de choix d'action.

Le rôle spécial des variables de comportement est également montré : ainsi que les variables motrices, elles sont considérées comme un genre spécial de variables d'état avec des valeurs connues.

6.3.2 Cognition

Le cadre proposé suggère de séparer le choix du comportement actuel des commandes motrices. Ceci est basé sur le principe des unités de base des mouvements coordonnés, appelées les synergies.

En particulier, le corps des vertébrés possède trop de degrés de liberté à commander par le cerveau. Un nombre relativement petit de mouvements coordonnés (certains très simples, d'autres plus complexes) sont choisis comme des mouvements naturels et ils sont appelés synergies. Ce répertoire de mouvements coordonnés constitue la base de tous les mouvements complexes : l'ordre d'exécution des synergies adéquates à chaque situation est choisi selon le comportement désiré.

Chapitre 7 : Sélection de l'attention

Dans ce chapitre, nous proposons de réduire le temps nécessaire pour calculer toutes les observations résultant des données sensorielles, en appliquant le concept de l'attention, où seulement les observations perceptuelles les plus pertinentes sont calculées et appliquées. La pertinence d'une observation est définie à partir de son utilité pour l'exécutions du comportement choisi pour le robot.

7.1 Programme bayésien

7.1.1 Description

Pour choisir des observations plus appropriées, un ensemble de variables C^t de choix d'attention est défini. Chaque variable dans cet ensemble correspond à un capteur physique ou à un canal de données d'entrée. Pour une variable, chaque valeur indique une méthode de transformation différente à appliquer aux données brutes.

Les variables C^t d'attention sont communes à tous les filtres élémentaires, et le modèle d'attention est défini en appliquant la fusion avec cohérence. Une variable de cohérence α^t est alors nécessaire, une variable dans chaque filtre élémentaire pour chaque

variable C^t.

Le programme bayésien présenté dans ce chapitre inclut un modèle de focalisation de l'attention pour décider quelles observations sont plus appropriées au pas de temps actuel. Ce modèle contient une dépendance à l'égard du comportement actuel.

$$\begin{aligned} & P(M^{0:t} S_i^{0:t} Z_i^{0:t} B^{0:t} C^{0:t} \lambda_i^{0:t} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i) \\ = & \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j C^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\ \times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i). \end{array} \right. \end{aligned} \tag{7}$$

7.1.2 Utilisation

L'utilisation de ce programme inclut une question de choix d'attention dans chaque filtre élémentaire. Les réponses sont combinées, et une valeur est tirée de la distribution globale résultante. Cette valeur indique comment traiter les perceptions et, par conséquent, à quelles observations des valeurs valides sont attribuées.

Une dépendance cyclique existe alors entre le choix de l'attention et le choix du comportement. Pour décider quelles données de la perception doivent être traitées, il serait souhaitable de savoir quel comportement est le plus approprié dans la situation actuelle. Cependant, pour décider quel comportement est le plus approprié, il est nécessaire d'avoir toute l'information disponible sur l'environnement, soit les observations.

Ce cycle crée un problème concernant l'ordre pour poser les questions d'attention et de comportement. L'approche bayésienne fournit une solution évidente à ce problème. La question de choix d'attention est posée avant la question de comportement et, quand les règles d'inférence sont appliquées, l'expression résultante indique quel comportement est prévu. Pour avoir la prévision globale de comportement au lieu de la prévision locale dans chaque filtre élémentaire, la question d'attention dans chaque filtre élémentaire inclut, parmi les variables recherchées, les variables de comportement. La question globale de l'attention, cependant, concerne seulement des variables de choix d'attention.

7.1.3 Exemple

L'exemple présenté dans le chapitres précédents est repris ici et, un modèle de focalisation de l'attention a été ajouté.

7.1.4 Hypothèses supplémentaires

- L'utilisation de la focalisation de l'attention réduit le temps nécessaire pour traiter les données des capteurs.

- L'attention vise à optimiser l'utilité de l'information disponible pour le comportement choisi au pas de temps actuel.
- L'attention détermine quelles observations sont disponibles pour mettre à jour les variables d'état.

7.2 Analyse de conditions

7.2.1 Complexité spatiale

L'inclusion de l'attention dans ce filtre augmente la quantité de mémoire nécessaire pour le modèle capteur et ajoute aussi un nouveau terme dans l'expression de mémoire lié au modèle de sélection d'attention.

Dans le cas des applications pratiques en robotique, ajouter le choix de comportement et les modèles de choix d'attention n'implique pas une augmentation de la façon dont les besoins de mémoire changent avec la taille de système.

7.2.2 Complexité temporelle

Le temps de calcul a été augmenté à cause de l'ajout de la question d'attention, et par conséquent l'expression de la complexité temporelle inclut la dépendance exponentielle avec les variables de comportement.

Cette dépendance devient négligeable si on considère les relations entre le nombre de variables souvent utilisées dans les applications pratiques en robotique.

7.2.3 Programmation

Des considérations sur le choix des variables d'attention sont présentées, ainsi que leur influence sur le modèles capteurs. Des commentaires sur la programmation du modèle d'attention sont aussi exposés.

7.3 Discussion

7.3.1 Robotique

Le déploiement de l'attention a été l'objet des recherches et des applications dans les systèmes artificiels. Ici, nous nous interessons aux modèles informatiques d'attention visuelle et à la vision active, à quelques applications en robotique, et à la façon dont ces modèles sont liés à notre proposition.

7.3.2 Cognition

Focaliser l'attention dans une partie de l'espace sensoriel et, en conséquence, négliger d'autres parties peut contribuer à diminuer l'effort cognitive de traiter tout les signaux venues de la perception.

Chapitre 8 : Expériences en robotique

Ce chapitre présente et discute les résultats de l'utilisation pratique des filtres proposés dans les chapitres précédents sur un robot réel.

8.1 Conditions expérimentales

Cette section présente les conditions expérimentales. D'abord, nous décrivons le comportement désiré du robot, comme cahier des charges pour l'expérience. Ensuite, le robot lui-même est détaillé : sa structure physique, ses interfaces avec le monde externe et les ressources informatiques embarquées. En conclusion, nous expliquons comment les ressources du robot sont employées pour mettre en application le comportement indiqué.

8.2 Programme bayésien

Des programmes bayésiens sont décrits, suivis de leur utilisation. Une fois que les valeurs des commandes motrices pour le pas de temps actuel sont connues, il est importante d'assurer leur exécution en toute sécurité, ce qui consiste à éviter les éventuels obstacles pendant l'exécution de ces commandes de moteur. La description de l'utilisation de filtres comprenant l'action d'éviter les obstacles est également présentée.

8.3 Résultats et analyse

Cette section montre le résultat de l'exécution du programme présenté ci-dessus sur le robot. Ce résultats sont analysées vis-à-vis du comportement spécifié et quelques remarques sur la mise en exécution du filtre proposé sont présentées.

Chapitre 9 : Conclusions et perspectives

9.1 Contributions

Notre contribution principale réside dans un cadre pour la programmation d'un robot mobile, basé sur des définitions mathématiques strictes et comportant le choix d'action et la perception sélective, en utilisant l'approche bayésienne.

Les contributions secondaires de ce travail concernent les solutions utilisées dans ce cadre, l'analyse et la discussion incluses dans ce document, et la méthode de programmation appliquée dans les expériences simulées et avec un robot réel.

9.2 Perspectives

Plusieurs travaux sont possibles à partir des propositions et résultats présentés, parmi lesquels : construction et utilisation de cartes de l'environnement, comparaison avec d'autres méthodes de sélection d'action, perception active, extension aux techniques hiérarchiques de sélection d'action.

Chapter 1

The Ultimate Question for Autonomous Sensory-Motor Systems

Key Points:

- **Autonomous sensory-motor systems continuously answer the ultimate question.**
- **Systems to answer the ultimate question should fulfil requirements.**
- **We propose several solutions of increasing complexity to the ultimate question.**

What similarities can be found between an animal and an autonomous mobile robot? Both can control their motor capabilities based on information acquired through dedicated channels. For an animal, motor capabilities are muscles and joints, and filtered information from the environment is acquired through sensors: eyes, nose, ears, skin, and several others. For a mobile robot, motor capabilities are mostly end effectors and mechanical motors, and information about the surroundings consists of data coming from sensors: proximeters, laser range sensor, bumpers, and others.

In this work, we assume that an autonomous sensory-motor system is continually answering the ultimate question: $P(M^t|z^{0:t} \wedge m^{0:t-1}\pi)$. This question can be put into words as: *What can I do next, knowing what I have seen and what I have done?*

Mathematically, the expression gives the probability distribution over the values of motor control variables M at time instant t, knowing the values of the observed variables Z from time instant 0 to time instant t, as well as the values of all motor controls exerted from time instant 0 to time instant $t-1$, under the previous knowledge about the robot, its task and the environment. All this previous knowledge is assembled and summarised by the variable π.

We elaborate, in the following chapters of this document, a framework that answers the ultimate question. This framework is applied to programming behaviours in a mobile robot. We claim that focus of attention and behaviour selection, combined to deal with uncertainty, contribute to building better performing robot control systems.

In this initial chapter, we begin our investigation of how to answer the ultimate question: the simplest way is to program it manually in a single table. This rudimentary solution is presented and analysed in relation to its feasibility in Section 1.1. Then, in Section 1.2 we list a set of basic requirements for any system expecting to solve the ultimate question in a real robot. Section 1.3 outlines the framework we propose for answering the ultimate question: a detailed presentation is given later in Chapter 7. From this first elementary solution to our final proposition, a succession of intermediate solutions is proposed in this document. Their structure is described in Section 1.3.

1.1 A Naive Answer to the Ultimate Question

This section discusses an extremely simple solution to the ultimate question. Through a careful analysis of this first basic solution, our objective is to deduce notions of more effective ways to answer the ultimate question in a robotic system.

Let us first compile all available information. The observed variables Z have known values from time instant 0 to time instant t. These values are indicated using small letters as $z^{0:t}$. The motor variables M also have known values, from time instant 0 to time instant $t-1$, and these values are denoted $m^{0:t-1}$. We assume all variables (including time) are assigned discrete values.

The simplest possible solution that results in the probability distribution over the motor variables M at time instant t is a table, manually filled with probabilities for all different possible values of observed and motor variables.

Because of the multi-dimensional aspect of such a table, a small example is illustrated in figure 1.1, showing one observation variable and one motor variable at time instant $t = 1$. This table contains probability values for expression $P(M^1|z^{0:1}m^0\pi)$, and it requires three indexes: z^0, z^1 and m^0. Each dimension in the table is then related to one possible value of each variable used as index. Each cell in the table stores the probability distribution over the motor variable values at time instant 1.

Each new variable adds one dimension, and as the time horizon extends, the dimensions of the table increase even faster: each additional time step adds as many dimensions as the total number of variables.

Utilisation of this table is very simple. At time step t, sensor values are read and processed resulting in values z^t. These values are used, together with the previous sensor values $z^{0:t-1}$ and the previous motor values $m^{0:t-1}$, as indexes in the table to access the probability distribution of the current motor variables M^t.

This solution is very simple and requires only one assumption (that variables, including time, are assigned discrete values), but is it really feasible?

To estimate how much data is stored in this table, let us assume that N_m is the number of variables for motor commands and that there are N_z observation variables. Let us also suppose that each variable is assigned discrete values with n different cases.

The proposed probability table at time instant t would then have as many dimensions as there are variables in the right part of the ultimate question: $N_z \times t$ plus

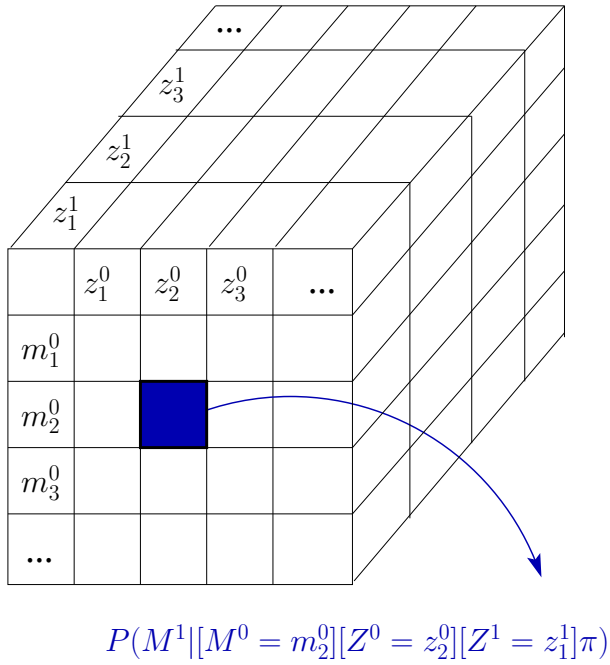


Figure 1.1: Naive solution for answering the ultimate question at time instant $t = 1$: one observation variable Z and one motor variable M.

$N_m \times (t-1)$. Each dimension has the same size, actually the number of cases for each variable, and consequently the table has $n^{N_z \times t + N_m \times (t-1)}$ cells. Each cell in the table contains not a value but a probability distribution over the variables M^t. Each cell therefore contains n^{N_m} probability values. In summary, to keep this table in memory, it is necessary to store $n^{(N_z+N_m)\times t}$ probability values.

If all variables have the same number of possible discrete values (for instance $n = 16$), and the present time is $t = 8$, the resulting table size is $2^{32\times(N_m+N_z)}$. Table 1.1 displays approximate numbers of probability values for the probabilities table as a function of the number of observation and motor variables.

As the memory requirements are unrealistic for any practical implementation, the time required to access the table cannot be evaluated. One access is necessary for each value of motor variable M^t to build the whole probability distribution, but with table dimensions increasing over time, it is not possible to estimate how long it takes to execute one access.

Another interesting point to consider when evaluating the feasibility of this proposal is the number of free parameters to be adjusted and how easily they can be adjusted. In this case, the parameters are the probability values that must be specified in the table. Thus, the number of parameters to be tuned is roughly as shown in table 1.1.

In addition to this incredibly huge number of parameters to adjust, manual programming of these values is rather difficult. Although it is possible to relate environment observations directly to motor commands in simple behaviours (obstacle avoidance or

	$N_z = 1$	$N_z = 2$	$N_z = 4$	$N_z = 16$
$N_m = 1$	10^{19}	10^{28}	10^{48}	10^{163}
$N_m = 2$	10^{28}	10^{38}	10^{57}	10^{173}
$N_m = 4$	10^{48}	10^{57}	10^{77}	10^{192}
$N_m = 16$	10^{163}	10^{173}	10^{192}	10^{308}

Table 1.1: Approximate numbers of probability values to be stored in relation to number of motor and observation variables, for $t = 8$ and $n = 16$: the expression is $2^{32\times(N_m+N_z)}$, where N_z is the number of observation variables and N_m is the number of motor variables. All variables are considered to have the same number of possible values, denoted n.

wall following, for example), when more complex situations are desired, it is very complicated to determine the necessary probability values.

Instead of determining these free parameters a priori, they can be learned from experimental data. However, the large number of parameters to be learned makes it hard to set up an adequate learning method for a robotic system.

For all the reasons discussed above, there is no evidence of practical use of a probability table in an autonomous system control unit. This initial discussion is useful to distinguish some of the main problems related to answering the ultimate question. These problems become the main motivations for developing the framework presented in this document.

Consequently, the analysis executed here allows us to outline the requirements for analysing other solutions. These requirements will be employed in the following chapters in this document to evaluate the proposed programming frameworks, and for comparisons between the different proposals.

1.2 Requirements

Obviously, this first proposed solution to the ultimate question is too simple and onerous: in all the aspects analysed above, it shows great handicaps.

The following chapters in this document propose solutions to the ultimate question, in a succession of incremental capabilities and cumulative hypotheses and simplifications. Each individual solution to the ultimate question is evaluated and compared with the proposals from previous chapters.

The above analysis identified requirements for examining proposed solutions, and these requirements are listed below, with brief analyses.

1.2.1 Space Complexity

The memory space required to store all data necessary for the solution must be within reasonable bounds. The probability table solution proposed above is an example of a solution that cannot be stored in a limited quantity of memory, except for very small problems of short time durations.

We can expect to make use of some hundreds of megabytes for a robotic system. In the following chapters, however, memory space will not be analysed numerically but in relation to the asymptotic curve describing how memory space increases with system parameters (as number of variables, for example). Estimation of this requirement is then called **space complexity analysis**.

1.2.2 Time Complexity

Reducing the amount of memory from the huge values shown above will increase the time necessary to obtain the results: instead of a table access, some calculations will be required, which means spending computational time.

Similarly to memory space analysis, the time consumed in calculations will be analysed mostly as the asymptotic curve showing how time increases with system size parameters, and estimation of this requirement is called **time complexity analysis**.

1.2.3 Programming

The programming task will certainly include modelling decisions and free parameters to be adjusted, regardless of the framework that is applied, the desired behaviour or the specific robot being programmed.

This requirement involves analysing how easy it is to model a specific behaviour in the proposed framework. This involves evaluating possible design choices, as well as the number and semantics of free parameters inherent in the framework.

In complex situations, learning methods can be applied to increase flexibility and to simplify the design of a robotic control system: some robotic tasks are hardly feasible if learning methods cannot be applied.

When analysing a proposed programming schema, its ability to benefit from learning methods is considered to be an important criterion of evaluation. Whenever possible, this evaluation must include an indication of which method to apply, and, if experimentation is relevant, it is also interesting to state the kind of experiment and evaluate the difficulty of setting it up.

1.3 Proposed Solution

Chapters 4 to 7 describe successive solutions to the ultimate question, with incremental complexity. This section presents briefly the solution proposed in Chapter 7, the last

and most complex of the presented solutions, which consists of an improved Bayes filter, including selective perception, behaviour selection and motor control.

It is clear that it is not possible to add the above capabilities employing only motor and observation variables; new variables are therefore added to the system as described below.

- State variables are internal variables representing useful characteristics of the robot and its surroundings.
- Behaviour variables indicate which kinds of action the robot can take.
- Attention variables are responsible for guiding selective perception.

Figure 1.2 shows the joint distribution equation of the proposed framework, with some indications of which terms in the equation effectively relate to the additional capabilities described above.

$$
\begin{aligned}
&P(M^{0:t} S^{0:t} Z^{0:t} B^{0:t} C^{0:t} \lambda^{0:t} \beta^{0:t} \alpha^{0:t} | \pi) \\
&= \left| \prod_{j=1}^{t} \left[\begin{array}{l}
\boxed{\prod_{i=1}^{N_i} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i)} \\
\boxed{\prod_{i=1}^{N_i} P(Z^j | S_i^j C^j \pi_i)} \\
\boxed{\times P(B^j | \pi) \prod_{i=1}^{N_i} P(\beta^j | B^j S_i^j B^{j-1} \pi_i)} \\
\boxed{\times P(C^j | \pi) \prod_{i=1}^{N_i} P(\alpha^j | C^j S_i^j B^j \pi_i)} \\
\boxed{\times P(M^j | \pi) \prod_{i=1}^{N_i} P(\lambda^j | M^j B^j S_i^j M^{j-1} \pi_i)}
\end{array} \right] \right. \\
&\quad \boxed{\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi)}
\end{aligned}
$$

Figure 1.2: Joint distribution of proposed framework. See text for explanation about color boxes.

A yellow rectangle indicates the part of the equation related to the initial conditions, or in other words, the values of all variables at the moment the system begins to work. The green rectangle marks the dynamics of states, and the blue rectangle indicates a sensor or observation model. The brown rectangle shows the motor command model, the red rectangle emphasises the action selection model and the orange rectangle marks the selective perception model.

Partial independence between different domains of interest in the environment can be exploited to reduce further the dimensionality of the problem while preserving coherence in system decisions. This leads to the definition of elementary filters, indicated by the magenta circles in the joint distribution in figure 1.2.

A sequence of questions is asked to this joint distribution at each time instant t, in the following specific order:

1. prediction question for each elementary filter;
2. attention question for each elementary filter;
3. fusion of attention questions and drawing attention values;
4. behaviour question for each elementary filter;
5. fusion of behaviour question and drawing behaviour values;
6. estimation at each elementary filter;
7. motor question at each elementary filter;
8. fusion of motor question and drawing motor command values.

1.4 Cognition

Living beings must also solve the ultimate question: what should be the contraction of muscles, the production of chemical products (for instance, pheromones, odours or toxins) and, eventually, the generation of electric fields knowing sensory inputs?

Required memory space and needed processing power computed in this chapter are of such high orders of magnitudes that they become an argument against behaviourist like approaches: the black box shall be open [1]. What are the information structures, memory management processes and computing mechanisms able to deal with such a huge, however incomplete, information flow within such strict time constraints?

This thesis proposes mathematical considerations that could be of interest as possible elements of answer to this fundamental question. Each chapter has a section dedicated to discussions about the cognitive implications of the mathematical hypotheses and simplifications presented there. We do not claim that these hypothesis are true from a cognitive point of view nor that they are a good model of the Central Nervous System because we have absolutely no biological proofs of their evidences. We rather discuss what would be their cognitive implications if they would be true.

[1] Behaviourist approaches focus on observable behaviour as inputs and outputs, but the internal processes relating the inputs to the outputs are not considered: this is the basic idea of the black box.

1.5 Contributions and Document Structure

The main contribution of this thesis is the proposal of a framework for programming autonomous robots, featuring Bayesian-based action selection and selective perception. The framework is founded on a succession of incremental hypotheses and assumptions, defined within the strict mathematical framework presented in Chapter 2, from which additional contributions emerge.

It is difficult to establish the direct relation between sensor observations and motor commands, except for simple behaviours. To simplify the process of writing this relation, a level of abstraction is introduced in Chapter 3 to model relevant environment features employing state variables.

The ultimate question considers all past observation and motor commands, which means memory space and time for calculation grow endlessly. One solution is to restrict past information to a reasonable depth. Chapter 4 applies the Markov assumption and stationarity to maintain memory space and calculation effort constant in time.

A secondary contribution in this chapter is the introduction of inference of motor commands in the Bayes filter, inside the same recursive loop of calculation.

Although the required memory and calculation effort thus become constant in time, they are still too big and are not limited in relation to system size. Any inherent independence between state variables can be exploited to reduce time complexity. Division of the state space into disjoint subsets and utilisation of the method of fusion with coherence are suggested in Chapter 5 to improve system scalability.

When the behaviour specified for the robot becomes more complex, direct association of environment features to motor commands is troublesome and ineffective. Insertion of another level of abstraction allows us to associate motor patterns to behaviour values. Chapter 6 deals with the addition of a behaviour coordination mechanism to the Bayes filter, which constitutes another contribution of this work.

The incoming flow of sensor data from a rich environment requires improved processing methods to extract desired features; the main consequence is long processing times for handling all the sensor data. Chapter 7 proposes to reduce this time by employing a mechanism of attention selection. The proposal of a Bayesian selective perception mechanism associated with action selection is an important contribution of this chapter.

Evidence of the practicability of the above proposed concepts when controlling a robot is given in Chapter 8. The method applied to develop this experimentation is an additional contribution.

Chapter 9 summarises contributions, main results and perspectives.

All chapters proposing solutions have a similar structure. Initially, the chapter proposal is presented by means of a Bayesian program; its presentation consists of description, utilisation, and an illustrative example. The proposal's main assumptions are then formally presented.

Next, the chapter proposal is examined: complexity in relation to space and time requirements is detailed, followed by analysis of the ease of programming.

Then, a general discussion regarding related works, constraints, significance and repercussions is given, considering robotics and cognition domains.

Chapters finish with a brief summary and open questions.

Secondary contributions of this work are the analysis of time and space complexities of Bayes filter descriptions and its utilisation for all the proposed filters, and the analysis of cognitive aspects of Bayes filter description and utilisation.

Chapter 2

Bayesian Framework for Dealing with Uncertainty

Key Points:

- **A rigorous mathematical method for dealing with uncertainty**

To answer the ultimate question effectively, we make use of concepts and methods to express our assumptions in a programming framework. In this chapter, we prepare the theoretical and practical tools employed in the following chapters.

We assumed in the previous chapter that we should deal with incompleteness and uncertainty by applying a probabilistic approach. Section 2.1 presents the theoretical approach used throughout this work to deal with uncertainty and incompleteness, called **Bayesian Programming**.

The ultimate question aims to establish a direct relation between motor and observed variables, but additional variables are necessary to express specified behaviours and relevant characteristics from the environment. These variables are called here **Internal Variables** ; they and their application in this thesis are described in Section 2.2.

As internal variables are added to the system, the complexity of expressing the relationships between all variables increases. Section 2.3 discusses conditional independence between variables as a method for reducing the intricacy of relations between variables.

The Bayesian programming approach allows propagation of the uncertainty of variable values. In particular situations, it may be desirable or necessary to stop uncertainty propagation for specific variables, which means making a decision over the variables' values. Section 2.4 presents briefly how the problem of decision making is addressed in this work.

2.1 Bayesian Approach for Reasoning and Expressing Uncertainty

To execute a previously specified behaviour, an autonomous sensory-motor system interacts with the environment.

Clearly, it is not possible to build an environment model able to predict all the possible events and configurations that could influence robot behaviour. Modelling implies incompleteness, particularly in a dynamic and unmodified environment.

Several methods have been proposed to deal with the diverse sources of uncertainty. A good review can be found in Parsons and Hunter [1998]. One possible approach is to include in the model a factor expressing its own uncertainty. The probabilistic approach is one possible representation of uncertainty, and Bayes theorem is the corresponding tool for reasoning under uncertainty.

Throughout this work, we apply the Bayesian robot programming framework, derived from the work of Jaynes [2003], which associates logic and probability in a precise and mathematically rigorous approach. The Bayesian robot programming framework was first proposed by Lebeltel [1999] in the context of mobile robot programming and later extended and employed for several applications [Bellot et al., 2004]: manipulator arm simulation and programming [Mekhnacha et al., 2001; Ramirez, 2003], automotive applications [Pradalier, 2004; Coué, 2003], and mapping [Diard, 2003].

2.1.1 Bayesian Robot Programming

Logic, Probability, and Reasoning under Uncertainty

Traditional logic's limitations are related to the distinction between *deductive* and *plausible* reasoning. Deductive reasoning allows reasoning under the logical consequence of events. Logical propositions are either true or false, implying true or false consequences [Jaynes, 2003].

In plausible reasoning, each logical proposition has an attached degree of plausibility, with bounds *not plausible*, meaning the proposition is false, and *completely plausible* indicating the proposition is true. When plausibility values are near limits (*not plausible* or *completely plausible*), plausible reasoning is similar to deductive reasoning.

The formal method introduced by Jaynes [2003] proposed applying probabilities as plausibility values, and this method supplies the rigid theoretical formalisation of Bayesian Robot Programming approach. Most relevant concepts and definitions of this formalisation are detailed below.

A **logical proposition** is a statement with a precise meaning and of definite logical type: either it is true or false. Logical propositions are denoted by small letters, and they can be combined, applying logical operators, to obtain new propositions.

As an example, if the proposition a means "*Robot battery requires charging*", and proposition b is "*Robot is placed for charge*", a new proposition $a \wedge b$ (the **conjunction** of a and

b, often noted ab) would have the meaning "*Robot battery requires charging and is placed for charge*". It is also possible to express the **disjunction** of propositions a and b, which results in a new proposition $a \vee b$ meaning "*Battery requires charging or Robot is placed for charge*". **Negation** of a proposition, as in $\neg a$, is another proposition meaning "*Battery does not require charging*".

Propositions can also mean that a variable has an associated value, as in proposition v meaning "*Robot battery voltage is 24 V*". We can say that, if the robot battery voltage is indicated by variable V (in capital letter), v means $[V = 24]$. This kind of proposition is used extensively when programming a robot and leads to the definition of a discrete variable.

A **discrete variable** V is then defined as a set of n logical propositions, one for each possible value in the domain $D_V = \{v_1, v_2, \ldots v_n\}$. These logical propositions are then mutually exclusive and exhaustive:

$$v_i \wedge v_j = \text{false, for } i \neq j;$$
$$v_1 \vee v_2 \vee \ldots v_n = \text{true.}$$

A variable can only be assigned values included in its domain, and only one value at a time.

A variable is always denoted by names beginning with capital letters, while the values in its domains are denoted by small letters. n denotes the cardinality of variable V, also indicated by $\lfloor V \rfloor$. As each possible value in the domain is associated with a logical proposition, there is no ambiguity in this notation: value v_j is equivalent to the proposition $V = v_j$.

The conjunction of two variables, as $X \wedge Y$ (or XY), is also a variable, but their disjunction is not a variable. Actually, the conjunction $X \wedge Y$ defines a set of propositions $x_i \wedge y_j$, which are mutually exclusive and exhaustive. On the other hand, the propositions in the set resulting from a disjunction $x_i \vee y_j$ do not have a mutually exclusive set of possible values, and consequently cannot be considered a variable.

To assign a plausibility value for a given proposition, it is necessary to have some background knowledge about the meaning and scope of the proposition. This background knowledge is called **preliminary knowledge**, and it is expressed by small Greek letters. They condition the assigned plausibility: $P(a|\pi)$ stands for *the plausibility that proposition a is true, knowing the preliminary knowledge summarised by π*.

The plausibility that the variable X has a value x_i is then expressed using a probability value $P([X = x_i]|\pi)$ or, identically, $P(x_i|\pi)$. It is also possible to write $P(X|\pi)$, as the plausibility of variable X being equal to each value in the domain D_X: in other words, the **probability distribution over values for variable** X.

Often, it is necessary to define the plausibility of a proposition conditioned by the knowledge of another proposition, as in $P(a|b \wedge \pi)$, which stands for *the plausibility of proposition a knowing that proposition b is true and given the preliminary knowledge π*. In this case, there is a **conditional dependence** between propositions a and b. It is often desired to know the probability distribution over the values of a discrete variable A conditioned to a known value of variable $[B = b_i]$, noted by $P(A|b_i \wedge \pi)$.

To reason using logical propositions with attached plausibility values, two postulates are sufficient for any computation: the conjunction and normalisation rules.

The **Conjunction rule** is derived from the well known Bayes rule, and it determines the probability of a conjunction of propositions:

$$\begin{aligned} P(a \wedge b|\pi) &= P(a|\pi) \times P(b|a \wedge \pi) \\ &= P(b|\pi) \times P(a|b \wedge \pi). \end{aligned} \tag{2.1}$$

The **Normalisation rule** defines the relation between the probability of one proposition and the probability of its negation: they must sum to one.

$$P(a|\pi) + P(\neg a|\pi) = 1. \tag{2.2}$$

From the above two rules, the conjunction and normalisation rules for variables are derived:

$$\begin{aligned} P(A \wedge B|\pi) &= P(A|\pi) \times P(B|A \wedge \pi) \\ &= P(B|\pi) \times P(A|B \wedge \pi). \end{aligned} \tag{2.3}$$

$$P(A|\pi) + P(\neg A|\pi) = 1. \tag{2.4}$$

Another rule that is very useful when reasoning with variables is the **marginalisation rule**, also derived from the conjunction and normalisation rules:

$$\sum_A P(A \wedge B|\pi) = P(B|\pi). \tag{2.5}$$

Rules presented above are limited to discrete variables, but similar reasoning is also possible to continuous variables, as presented in Mekhnacha et al. [2001] and Coué [2003].

Bayesian Programming

The above definitions and concepts allow us to use a set of probability distributions to describe the available knowledge about a set of variables. Based on this description, it is possible to calculate the probability distribution over a subset of variables, knowing the values of some or all other variables.

It is not easy, however, to organise the available information and be sure of its mathematical validity regarding the postulates and assumptions.

Bayesian programming is a unifying framework for formally describing probabilistic models, as demonstrated by Diard [2003] and Diard et al. [2003]. Below, the Bayesian programming framework is presented briefly. For more details about definitions and concepts involving Bayesian programming see Bessiere and Team [2003]. A broader discussion regarding the basic assumptions and hypotheses of this framework can be found in Bessière et al. [1999a,b].

The main object of Bayesian programming is the Bayesian program. It consists of a *description* of previous knowledge and experimental data, and a *utilisation* of this description.

The Bayesian program is a describing tool, and it assumes nothing about inference methods: whenever simplifications or approximations are relevant or necessary for efficient inference, they are described separately.

Description

The description part of a Bayesian program is a detailed report of relevant information available about the problem or situation under study.

Some parts of the description are fixed a priori, but others can result from learning or automatic tuning processes. That is why the description is composed of two parts: the first is called the *specification* and consists of the preliminary knowledge (π); the second comprises the *experimental data* and is related to data δ acquired experimentally for use in learning or tuning. Defining a Bayesian program consists of describing the preliminary knowledge (π) and the experimental data involved (δ).

The purpose of a description is to specify an effective method to compute a joint distribution on a set of relevant variables $\{X_1, X_2, \ldots, X_n\}$, given a set of experimental data δ and the preliminary knowledge π.

In the specification phase of the description, it is necessary:

- to define a set of relevant variables $\{X_1, X_2, \ldots, X_n\}$, on which the joint distribution must be defined;

- to decompose the joint distribution into simpler terms, following the conjunction rule. Given a partition of $\{X_1, X_2, \ldots, X_n\}$ into k subsets, with variables $\{L_1, L_2, \ldots, L_k\}$ each being the conjunction of the variables in the corresponding subset, the conjunction rule leads to the following decomposition of the joint distribution:

$$P(L_1\ L_2\ \ldots\ L_k \mid \delta\ \pi) = \left| \begin{array}{l} P(L_1 \mid \delta\ \pi) \\ \times P(L_2 \mid L_1\ \delta\ \pi) \\ \times P(L_3 \mid L_2\ L_1\ \delta\ \pi) \\ \times \ldots P(L_k \mid L_{k-1}\ \ldots L_2\ L_1\ \delta\ \pi). \end{array} \right. \tag{2.6}$$

  Conditional independence hypotheses can allow further simplifications. A conditional hypothesis for variable L_i is defined by choosing some variables X_j from among the variables appearing in conjunction $L_{i-1}\ \ldots L_2\ L_1$, calling R_i the conjunction of these chosen variables and setting:

$$P(L_i \mid L_{i-1}\ \ldots L_2\ L_1\ \delta\ \pi) = P(L_i \mid R_i\ \delta\ \pi). \tag{2.7}$$

The joint distribution then obtained is:

$$P(L_1\ L_2\ \ldots\ L_k \mid \delta\ \pi) = \left| \begin{array}{l} P(L_1 \mid \delta\ \pi) \\ \times P(L_2 \mid R_2\ \delta\ \pi) \\ \times P(L_3 \mid R_3\ \delta\ \pi) \\ \times \ldots P(L_k \mid R_k\ \delta\ \pi). \end{array} \right. \tag{2.8}$$

Such a simplification of the joint distribution as a product of simpler components is called a decomposition.

- Define the forms for each term $P(L_i|R_i\ \delta\ \pi)$ in the decomposition, i.e. each term is associated with either a parametric form (as a function $f_\mu(L_i)$) or with another Bayesian Program. In general, μ is a vector of parameters that may depend on R_i or δ or both.

$$\text{Program}\left\{\begin{array}{l} \text{Description}\left\{\begin{array}{l} \text{Specification}\left\{\begin{array}{l} \text{Appropriate Variables} \\ \quad X_1 \in D_{X_1}, \lfloor X_1 \rfloor \\ \quad X_2 \in D_{X_2}, \lfloor X_2 \rfloor \\ \quad \ldots \\ \quad X_n \in D_{X_n}, \lfloor X_n \rfloor \\ \text{Decomposition} \\ P(X_1\ X_2\ \ldots\ X_n \mid \pi) = \\ \quad P(L_1 \mid \delta\ \pi) P(L_2 \mid R_2\ \delta\ \pi) P(L_3 \mid R_3\ \delta\ \pi) \ldots P(L_k \mid R_k\ \delta\ \pi) \\ \text{Parametric Forms} \\ \quad P(L_1 \mid \delta\ \pi) = \ldots \\ \quad P(L_2 \mid R_2\ \delta\ \pi) = \ldots \\ \quad P(L_3 \mid R_3\ \delta\ \pi) = \ldots \\ \quad \ldots \\ \quad P(L_k \mid R_k\ \delta\ \pi) = \ldots \end{array}\right. \\ \text{Identification :} \\ \quad \text{A Priori or Learning Method} \end{array}\right. \\ \text{Question :} \\ \quad P(\text{Searched} \mid \text{Known}\ \delta\ \pi) \end{array}\right.$$

Figure 2.1: Bayesian Program General Structure

Utilisation

Given a description $P(X_1\ X_2\ \ldots\ X_n \mid \delta\ \pi)$, a question is asked by partitioning the variables $\{X_1, X_2, \ldots, X_n\}$ into three sets of variables: *Searched*, *Known* and *Unknown*. A question is defined as the distribution:

$$P(\text{Searched} \mid \text{Known}\ \delta\ \pi).$$

To answer this question, the following general inference is used, based on the conjunction and normalisation rules described in Equations 2.1, 2.3 and 2.5.

$$P(\text{Searched} \mid \text{Known } \delta\ \pi) = \frac{P(\text{Searched Known} \mid \delta\ \pi)}{P(\text{Known}|\delta\ \pi)}$$
$$= \frac{\sum_{\text{Unknown}} P(\text{Searched Unknown Known} \mid \delta\ \pi)}{\sum_{\text{Unknown,Searched}} P(\text{Searched Unknown Known} \mid \delta\ \pi)}$$

Figure 2.1 illustrates the general structure of a Bayesian Program. In summary, a Bayesian program consists of a description and a utilisation. The description phase is built from the specification, where the preliminary knowledge of the programmer is declared, and from identification, where experimental data are taken into account. The specification of preliminary knowledge contains the list of relevant variables, the decomposition of the joint distribution into a product of simpler terms, and the definition of the parametric form for each term in the decomposition.

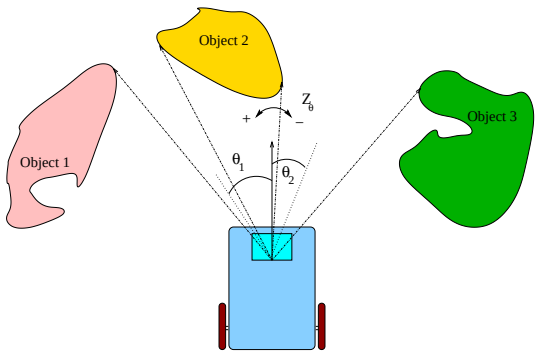


Figure 2.2: Robot and farthest direction

Example

Consider a two-wheeled robot equipped with a laser range sensor, which detects any obstacle in front of and around the robot, as illustrated in Figure 2.2. By controlling the speeds of both wheels separately, it is possible to change the robot's forward progress as well as its rotation speed.

It is desired to make the robot follow the farthest direction detected by the laser range sensor. In other words, the robot should advance in the direction where the detected obstacles are most distant. The solution proposed here is adapted from Mansard [2003] and Mansard et al. [2005].

The laser range sensor works by scanning 180 degrees in a plane: the time of flight of the laser beam indicates whether an obstacle is detected at a specific angle, and its distance. At the end of each scan, 361 measures are available (the distance of a detected obstacle for each 0.5 degree). If no obstacle is detected, the distance is set to the highest possible value (in this case, eight metres).

A preprocessing algorithm is applied to these data, resulting in the direction with the greatest distance without an obstacle. Variable Z_θ indicates the direction angle in relation to the robot's forward direction, while variable Z_{dist} is the distance in this direction. To avoid spurious data, or directions where the robot cannot pass, a minimum opening angle of 15 degrees in this direction is required. Figure 2.2 shows two possible directions to follow: θ_1 is actually too narrow, and so direction θ_2 is chosen.

Program
- Description
  - Specification
    - Appropriate Variables
      - $Z_\theta \in \{-5, 5\}, \lfloor Z_\theta \rfloor = 11$
      - $Z_{dist} \in \{0, 5\}, \lfloor Z_{dist} \rfloor = 6$
      - $M_{trans} \in \{0, 3\}, \lfloor M_{trans} \rfloor = 4$
      - $M_{rot} \in \{-2, 2\}, \lfloor M_{rot} \rfloor = 5$
    - Decomposition
      - $P(Z_\theta\ Z_{dist}\ M_{trans}\ M_{rot} \mid \pi) =$
        $P(Z_\theta \mid \pi) P(Z_{dist} \mid \pi) P(M_{rot} \mid Z_{dist}\ Z_\theta\ \pi) P(M_{trans} \mid Z_{dist}\ \pi)$
    - Parametric Forms
      - $P(Z_\theta \mid \pi) =$ Uniform;
      - $P(Z_{dist} \mid \pi) =$ Uniform;
      - $P(M_{rot} \mid Z_{dist}\ Z_\theta\ \pi) = G_{\mu_{rot}(Z_\theta);\sigma_{rot}(Z_{dist})}(M_{rot})$;
      - $P(M_{trans} \mid Z_{dist}\ \pi) = G_{\mu_{trans}(Z_{dist});\sigma_{trans}(Z_{dist})}(M_{trans})$.
  - Identification :
    - A Priori
- Question :
  - $P(M_{trans}\ M_{rot} \mid z_\theta\ z_{dist}\ \pi)$

Figure 2.3: Bayesian Program for following farthest direction

The program must control the speed of both the robot's wheels to make the robot follow the direction with the biggest distance. Instead of controlling the left and right wheels (variables M_{left} and M_{right}, respectively), it is more intuitive to control translation speed (forward progress, variable M_{trans}) and rotation speed (turning rate, variable M_{rot}). The transformation is defined by equations:

$$\begin{aligned} M_{left} &= M_{trans} + M_{rot}, \\ M_{right} &= M_{trans} - M_{rot}. \end{aligned} \tag{2.9}$$

The sign for the rotation speed M_{rot} is the same as that for Z_θ, as shown in Figure 2.2: positive when turning left, negative when turning right.

The above description allows us to define the set of pertinent variables, as shown in the first part of the Bayesian program shown in Figure 2.3.

The second part of the Bayesian program refers to the decomposition of the joint distribution over the relevant variables. The opening can be in any direction around the robot and at any distance, so the first two terms in the decomposition show that the values of Z_θ and Z_{dist} are independent of each other and of other variables.

Knowing the direction and distance of the opening, it is possible to define a probability distribution over the rotation speed. This is expressed by the third term in the

decomposition.

$$P(M_{rot} \mid Z_{dist}\ Z_\theta\ \pi).$$

Depending on the depth of this opening (the distance to the nearest obstacle in this direction), the robot's speed will vary. The last term in the decomposition declares this knowledge in the form: knowing the distance to the obstacle (in the opening direction) it is possible to define the probability distribution over the translation speed.

$$P(M_{trans} \mid Z_{dist}\ \pi).$$

As no a priori information about the direction or distance of this opening is available, the parametric forms for Z_θ and Z_{dist} are chosen to be uniform.

The parametric forms chosen for the remaining terms are normal distributions. For the translation speed term, the mean and standard deviation are defined as functions of the distance. For the rotation speed, the mean value depends on the direction angle, and the standard deviation depends on the distance. No learning is applied, and the functions for the parameters (the above means and standard deviations) are given a priori, as shown in Figures 2.4 and 2.5.

The curves shown in Figure 2.4 are related to the definitions of the translation speed terms. $P(M_{trans} \mid Z_{dist}\ \pi)$ consists of a family of normal distribution curves (bottom part of figure), with means and standard deviations defined according to sigmoid curves (top part of figure). It can be seen that the robot can only speed up if the distance is large: the mean speed increases with distance. The standard deviation is small when the distance is small: in this situation, it is important to be sure that a mean translation speed near zero will be imposed on the motor command. As the distance increases, the standard deviation increases, and thus also the uncertainty around the mean value (bigger translation speeds are possible in this situation). Following this reasoning, the resulting distribution curves are narrow when the distance is small (the robot must stop, or advance very slowly), but they became near uniform if the distance is large (the robot can advance slowly or quickly, no constraints are imposed).

In the bottom part of the figure 2.4, it can be seen that for each value of the distance variable, one curve is associated to the probability distribution over the translation speed value. Once the value of Z_{dist} is known, one value for mean and one value for standard deviation (according to curves in the top of figure) define the shape of the curve for $P(M_{trans})$. For example, if Z_{dist} is equal to 1 (object in the opening direction is near), mean value is very small (according to top left curve) and standard deviation is also very small (as can be seen in top right curve). Probability distribution for translation speed is then defined according to pink curve, where we can see that the small values of translation speed are more probable than the big values of translation speed.

Figure 2.5 shows the curves related to rotation speed. $P(M_{rot} \mid Z_\theta\ Z_{dist}\ \pi)$ is composed of a family of normal distributions for several values of distance and direction angle. As can be seen in the mean function curve, the rotation speed has the same signal for direction, but its absolute value does not increase linearly with angle, to avoid

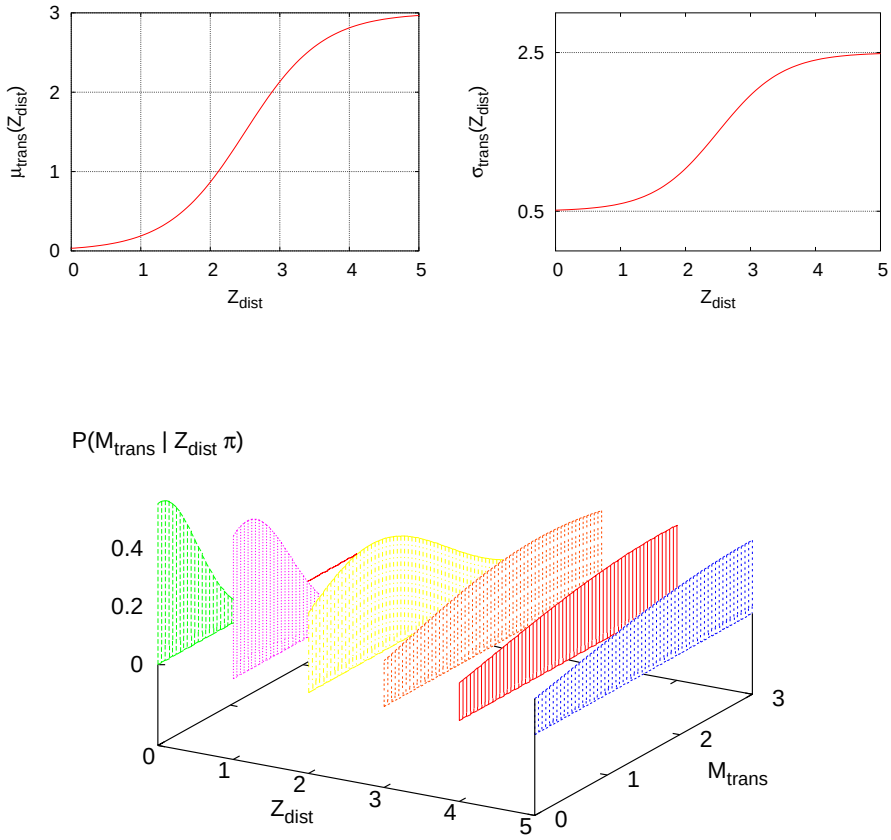


Figure 2.4: Farthest direction program. Top: curves for mean and standard deviation. Bottom: family of normal distributions for translation speed.

abrupt turns. The standard deviation curve is the same as for the translation speed: it is more certain in situations when the robot must not move. To illustrate the variation of rotation speed with direction angle, a fixed value of distance is chosen (and this therefore fixes the standard deviation). The curves then indicate which direction to follow according to the direction angle.

To use this Bayesian program, we ask it the question $P(M_{trans}\ M_{rot} \mid z_\theta\ z_{dist}\ \pi)$. In this simple example, the answer to this question is:

$$P(M_{trans}\ M_{rot} \mid z_\theta\ z_{dist}\ \pi) \propto P(M_{rot} \mid z_{dist}\ z_\theta\ \pi) \times P(M_{trans} \mid z_{dist}\ \pi). \tag{2.10}$$

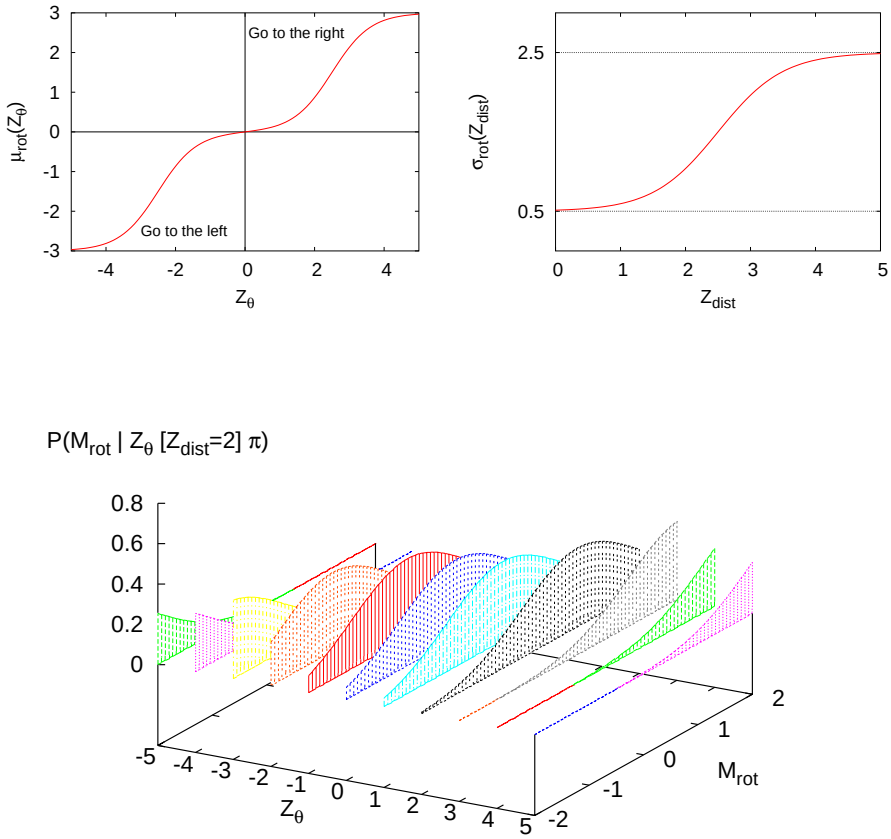


Figure 2.5: Farthest direction program. Top: curves for mean and standard deviation. Bottom: family of normal distributions for rotation speed when distance $Z_{dist} = 2$.

2.2 Internal Variables

When designing a system to control a robot and answer the ultimate question, input variables are associated with perceptions (in other words, measures from sensors), as output variables are related to actuators. Except for very simple behaviours, it is hardly possible to establish a direct relationship between input and output variables for a robot control system.

As an additional complication, sensors usually are not able to supply all the information necessary about the environment: processing as well as fusion of information from several sensors is often necessary.

Some level of abstraction based on the sensor measures is unavoidable, and it is usually achieved by employing internal variables, called states.

Internal variables fulfil several functions, for example they describe environment features, reason and deduce internal decisions, and express behaviour-inherent quant-

ities.

In Chapter 3, utilisation of internal variables to describe system conditions is described, and further functions of internal variables can be identified in Chapters 4, 5, 6 and 7.

2.3 Conditional Independence of Variables

As we saw in the previous section, to model complex behaviour, internal variables are essential. As the number of variables in the control system increases (observation, motor and internal variables), it becomes critical to consider variable dependencies.

In a complete model, it is assumed that all variables depend on each other, but in practical situations, especially with a very large number of variables, considering all possible dependencies between variables implies infeasible complexity.

In these situations, conditional independence can be useful to reduce the problem complexity. In reality, two variables can be independent, but conditioned by a third variable.

As shown in Section 2.1.1, conditional independence between variable sets L_1 and L_3 knowing variable set L_2, for example, can bring interesting simplifications. Term $P(L_3 \mid L_2\ L_1\ \delta\ \pi)$ can then change to $P(L_3 \mid L_2\ \delta\ \pi)$:

$$P(L_1\ L_2\ \ldots\ L_n \mid \delta\ \pi) = \left| \begin{array}{l} P(L_1 \mid \delta\ \pi) \\ \times P(L_2 \mid L_1\ \delta\ \pi) \\ \times P(L_3 \mid L_2\ \delta\ \pi) \\ \times \ldots P(L_k \mid L_{k-1}\ \ldots L_2\ L_1\ \delta\ \pi) \end{array} \right. \tag{2.11}$$

A practical example is when several sensors, identical or not, measure the same phenomenon. A basic assumption widely considered is that sensor readings depend only on the phenomenon being measured, and no dependence between the sensor readings is considered once the phenomenon is given. Whenever it is possible and reasonable, conditional independence will be applied to reduce intricacy in joint distributions.

2.4 Decision Making

Decision making is a research field in itself, and it is not possible to present here an extensive and ordered summary. We aim to present decision making as it is applied in our work: a stop in the propagation of uncertainty.

Using the probabilistic approach to handle uncertainty allows dealing with variables without an assigned value. In reality, a probability distribution over the variable values is attributed to each variable. Instead of using a variable value in the calculations, the associated probability distribution is employed, and this is called propagation of uncertainty.

One of the drawbacks of uncertainty propagation is the large increase in required calculations: rather than a single value, a set of values is applied everywhere the variable is used.

In a system like the one proposed here to control a robot, numerous variables are used, and if uncertainty propagation is essential for some of them, for others it is not important or can eventually be inconvenient. For these, a decision can be taken in relation to their values, stopping the uncertainty propagation process. These variables now no longer have probability distributions but assigned values.

The choice of the value to be assigned is often based on optimisation (cost or risk minimisation, and benefits maximisation). Here, the decision is taken by assuming that the probability distribution over the variable value contains all necessary information for taking the decision. It is then possible to select the most probable value or draw a value according to the whole distribution.

2.5 Summary and Open Questions

This chapter presented the methods applied throughout this document. The Bayesian Programming approach was presented as the theoretical unifying tool for describing models applying Bayesian inference.

Other useful resources such as utilisation of internal variables and expressing conditioned independence between variables were also discussed.

Propagation of uncertainty regarding variables values can be a blessing or a curse, depending on the situation. We have presented our approach of taking decisions on variable values when it is desired to stop propagation of uncertainty for these variables.

Chapter 3

State of the World

Key Points:

- **Addition of State variables reduces difficulties in programming.**
- **State variables uncouple motor and observation variables.**

When trying to solve the ultimate question by the solution with the simplest structure, namely a table with probability values as proposed in Chapter 1, it is clear that it is not at all suitable from a practical point of view. One of the problems is the difficulty in expressing a direct association between observation values and motor commands.

This chapter deals with the addition of state variables, which are widely employed in robotics. They concern environment features most relevant to the execution of specified behaviour, and they are also related to robot dynamic features (such as position, direction, and speed).

The use of state variables simplifies the utilisation of sensor signals and the definition of motor commands, uncoupling observation and motor command variables.

3.1 Bayesian Program

We present in this chapter a Bayesian program for solving the ultimate question at a given time instant j. The relation of variables at the present time with variables at past time instants is not considered here. Our goal in applying such a simplification is to show the utility of state variables when solving the ultimate question. We postpone dealing with problems associated with time dependence to the next chapter.

This Bayesian program shows how to add state variables to uncouple sensor observations and motor commands. The Bayes fusion schema is also presented as a way of improving the computation of state variable probability distributions by combining evidence from several sensor observations.

3.1.1 Description

The description of this Bayesian program is restricted to relevant variables and decomposition, as parametric terms are very problem dependent. An illustrative example is shown in Section 3.1.3 where parametric forms are discussed in the context of this example.

Additional Relevant Variables

In addition to the set of observation variables Z and the set of motor variables M, this program includes state variables S. State variables are chosen as the minimum set of relevant variables related to the robot, to the specified task and to the most significant features in the robot environment. All these variables are considered at time instant j.

Additionally, Z^j (the set of observation variables at time instant j) is considered as being composed of N_z observation variables, so that:

$$Z^j = \{{}_1Z^j, {}_2Z^j, \ldots {}_{N_z}Z^j\}, \forall j. \tag{3.1}$$

Decomposition

The joint distribution over the set of relevant variables listed above is defined as:

$$\begin{aligned} & P(S^j {}_1Z^j {}_2Z^j \ldots {}_{N_z}Z^j M^j | \pi_s) \\ = & \left| \begin{array}{l} P(S^j|\pi_s) \\ \prod_{k=1}^{N_z} [P({}_kZ^j|S^j\pi_s)] \\ \times P(M^j|S^j\pi_s). \end{array} \right. \end{aligned} \tag{3.2}$$

The first term in this joint distribution shows a priori information about state variables.

The second line in the equation is the product of terms related to observation variables. These terms are called sensor models, and they establish the relation between sensor observations and the environment state. The product of terms indicates that sensor observations are considered independent of each other if state variables are known.

The last term in the equation is the definition of how motor variables change with state values.

It can be noticed that state variables dissociate observation variables and motor commands. Observations are linked to states in the sensor model, and motor commands are functions of states in the motor model.

The sensor model is written based on sensor supplier information. A sensor measure is related to a phenomenon, which is associated to a state variable. If the phenomenon and its associated state variable are known, the sensor measure can be calculated.

When several sensors are related to the same state variable (in a redundant or complementary way), it is assumed that the state variable value is sufficient to determine

the values of all sensors readings. Sensor observations are then considered independent given the state variables, and this assumption implies the product of the sensor model in the above joint distribution.

The motor model defines how robot activity (defined by motor commands) depends on the environment and the specified task (described by the state variables).

3.1.2 Utilisation

This program can be used to estimate state variables from sensor measures. In this case, we put the following question to the joint distribution, with state variables in the left part of the question (searched variables, in capital letters) and the set of sensor measures in the right part (known variables, in small letters). Developing the question as shown in equation **??**, we obtain:

$$\begin{aligned} & P(S^j|_1 z^j{}_2 z^j \ldots_{N_z} z^j \pi_s) \\ & \quad \propto \quad P(S^j|\pi_s) \prod_{k=1}^{N_z} \left[P({}_k z^j | S^j \pi_s) \right] \quad . \end{aligned} \tag{3.3}$$

Examining the expression above, we can see that the state estimation is given by a combination of a priori information and fusion of individual sensor estimations.

Another utilisation of this program is to find the motor command probability distributions, given the sensor measures. For this question, the sought variables are the set of motor commands, the known variables are the set of sensor measures, and the state variables are unknown. In this case, the question asked of the joint distribution is:

$$\begin{aligned} & P(M^j|_1 z^j{}_2 z^j \ldots_{N_z} z^j \pi_s) \\ & \quad \propto \sum_{S^j} \left[\begin{array}{l} P(S^j|\pi_s) \prod_{k=1}^{N_z} \left[P({}_k z^j | S^j \pi_s) \right] \\ \times P(M^j | S^j \pi_s) \end{array} \right]. \end{aligned} \tag{3.4}$$

The sum over state space results from the fact that state variables have no attributed values, as the Unknown variables in equation **??**. State variables estimation as shown in equation 3.3 is present in the motor command question, so it is possible to rewrite above expression as:

$$\begin{aligned} & P(M^j|_1 z^j{}_2 z^j \ldots_{N_z} z^j \pi_s) \\ & \quad \propto \sum_{S^j} \left[\begin{array}{l} P(S^j|_1 z^j{}_2 z^j \ldots_{N_z} z^j \pi_s) \\ \times P(M^j | S^j \pi_s) \end{array} \right]. \end{aligned} \tag{3.5}$$

The probabilities of motor commands are calculated by weighting the motor model by the state estimates.

3.1.3 Illustration Example

To illustrate the Bayesian program described above, we consider a robot with eight ambient light sensors and two wheels. We want this robot to execute a phototaxy behaviour [Lebeltel, 1999]. Each sensor has a narrow angle of view, and to cover all the space around the robot, they are spaced around the circumference, as shown in Figure 3.1. In some directions, only one sensor is able to detect the existence of a source of light, but at other angles, two or more sensors can detect it.

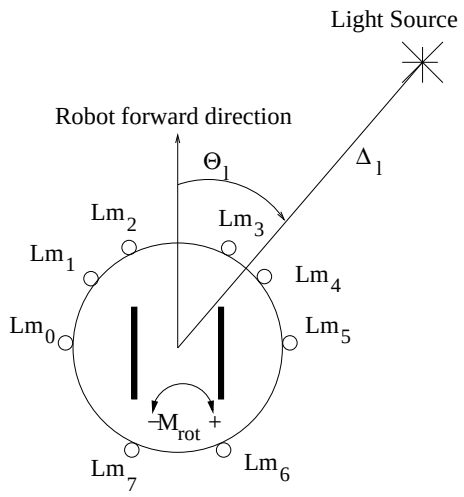


Figure 3.1: Top view of robot and position of ambient light sensors [Lebeltel, 1999].

Bayesian Program

The position of the light source is indicated by two state variables: Θ_l is the light source direction in relation to the robot's forward direction, and Δ_l is the distance to the light source. The direction variable has 36 different possible values, from –180 degrees to +170 degrees. The distance variable can take values from 0 cm (light very near) to 25 cm (light far away), with 26 different possible values.

The observation variables are the set of eight sensor measures (from Lm_0 to Lm_7), and these measures can be determined knowing the light source position in relation to robot. Each sensor variable can assume one of 512 possible values: from 0 (bright light) to 511 (dim light).

Phototaxy behaviour is equivalent to making the robot follow the direction of the light. Robot movement is commanded by controlling right and left wheel speeds, but it is possible to control the robot in terms of translation and rotation speeds. For the phototaxy behaviour, the translation speed is considered constant, and the robot direction is controlled by the speed of rotation associated with variable M_{rot}. This variable has 21 different possible values, from –10 to +10, where positive speeds correspond to the robot's turning right.

The joint distribution for this example is then:

$$\begin{aligned} & P({\Theta_l}^j {\Delta_l}^j Lm_0^j Lm_1^j ... Lm_7^j {M_{rot}}^j | \pi_e) \\ = & \left| \begin{array}{l} P({\Theta_l}^j {\Delta_l}^j | \pi_e) \\ \prod_{k=0}^{7} \left[P({Lm_k}^j | {\Theta_l}^j {\Delta_l}^j \pi_e) \right] \\ \times P({M_{rot}}^j | {\Theta_l}^j \pi_e). \end{array} \right. \end{aligned} \tag{3.6}$$

The first term in this joint distribution expresses the a priori knowledge about the position of a light source. We assume that the light source can be located at any distance and direction, so a uniform distribution is attributed to this term.

The second term in the joint distribution is the fusion of sensor models. Sensor supplier documentation describes the sensor working properties, and based on this description it is possible to write the model of one sensor: the probability distribution over the measure of one specific sensor Lm_i given the position of a light source.

The position of the light source is sufficient to determine each sensor reading separately, so these sensors have independent measures given the position of the light source. As we have eight sensors installed in the robot, we have eight sensor model terms. These eight models are similar, but as each sensor has a different position around the robot, their relation to the light's direction changes accordingly in each model. More details about the parametric forms chosen for this sensor can be found in Chapter 5 of Lebeltel [1999].

In the last line of the joint distribution, the motor command model defines the robot rotation speed M_{rot}, given the direction of the light source Θ_l. It is very simple to write the motor model, as there is a simple relation between the direction of the light and the direction in which the robot will turn: if Θ_l indicates the light is on the robot's right side, the robot must turn right and positive M_{rot} values have the highest probability values. The parametric form for this term is chosen to be a normal curve; its mean changes linearly with the direction of light, and its standard deviation is fixed.

$$P({M_{rot}}^j | {\Theta_l}^j \pi_e) = \mathcal{G}({M_{rot}}^j), \mu = {\Theta_l}^j / 1.8, \sigma_e. \tag{3.7}$$

This program can be used to estimate the position of the light source and to control the robot. When estimating the position of the light source, it is desired to know the distribution over states (the left part of the question) given the sensor measures (in the right part of the question):

$$\begin{aligned} & P({\Theta_l}^j {\Delta_l}^j | lm_0^j lm_1^j ... lm_7^j \pi_e) \\ & \propto \ \prod_{k=0}^{7} \left[P({lm_k}^j | {\Theta_l}^j {\Delta_l}^j \pi_e) \right] \ . \end{aligned} \tag{3.8}$$

The expression solving this question shows that the product of the sensor model terms is a fusion of sensor outputs to determine the probability distribution over states.

Another utilisation of this Bayesian program is to calculate the probability distribution over the motor variables (rotation speed) to control the robot. In this question, the left part is composed of the motor variables and the right part is composed of the sensor variables. This question is a simplified version of the ultimate question for the phototaxy behaviour at a time instant j.

$$\begin{aligned} & P({M_{rot}}^j | lm_0^j lm_1^j ... lm_7^j \pi_e) \\ & \propto \sum_{{\Theta_l}^j, {\Delta_l}^j} \left[\begin{array}{l} \prod_{k=0}^{7} \left[P({lm_k}^j | {\Theta_l}^j {\Delta_l}^j \pi_e) \right] \\ \times P({M_{rot}}^j | {\Theta_l}^j \pi_e) \end{array} \right]. \end{aligned} \tag{3.9}$$

The expression shows a summation over state variables ${\Theta_l}^j$ and ${\Delta_l}^j$. This is necessary, as the values of these variables are not known: they are actually estimated by sensor model fusion, but they do not have assigned values. The motor model is then weighted by this estimation.

Simulation Results

Figures 3.2 and 3.3 show some simulation results for the above program. First, the state estimation question is illustrated in Figure 3.2. The light source is positioned in front of the robot, slightly to the right, and at a medium distance. The real value for ${\Theta_l}^j$ is 10 and for ${\Delta_l}^j$ is 10.

Sensors at the front and right side of the robot (Lm_2, Lm_3 and Lm_4) give uncertain information about where the light might be located: the light source is within their angle of view. All other sensors generally indicate directions where the light cannot be located: this happens because the light source is outside their angle of view.

Figure 3.3 shows the result of the motor question: the probability distribution of the rotation speed variable when the light source is located as described above. We can see that the higher probability values are for no rotation and for a slight rotation towards the right direction, which is consistent with the position of the light.

3.1.4 Additional Main Assumptions

- State variables uncouple sensor observation and motor command variables.

  Relevant features of the environment and robot task are described by state variables. For this reason, it is easier to write motor commands related to states than to sensor observations. Sensor observations can also be interpreted and combined to supply evidence on relevant aspects of the world, such as those described by state variables.

  It is assumed that the set of state variables is sufficient to dissociate observations and motor variables.

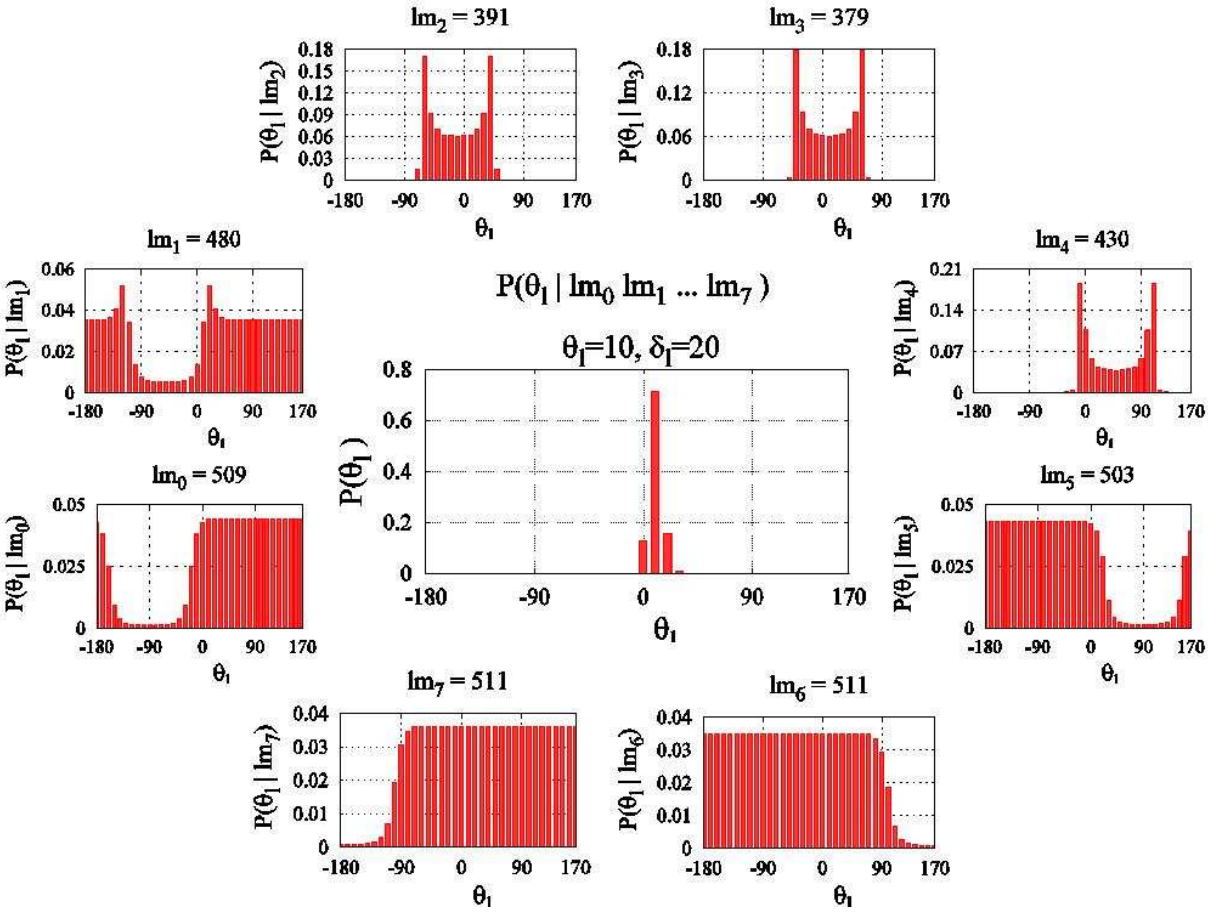


Figure 3.2: Fusion of light sensors for light direction estimation. All curves show probability distributions over values of the variables Θ_l. The central curve is the result of fusion of the eight sensors. The sensor curves are disposed according to the positions of the sensors in the robot, and the value of each sensor measure is indicated on the top of the estimation.

- Observations consist of measures from sensors, which can be modelled as functions of the state variables.

  Knowing the state of the world and of the robot, it is possible to infer sensor measures: this is the role of the sensor model. It is assumed that state variables are sufficient to indicate the most probable sensor readings.

- Sensor observations are independent when internal state variables are known.

  Whenever several observation variables are related to the same set of state variables, it is assumed that these observations are independent of each other if the set of state variables is known.

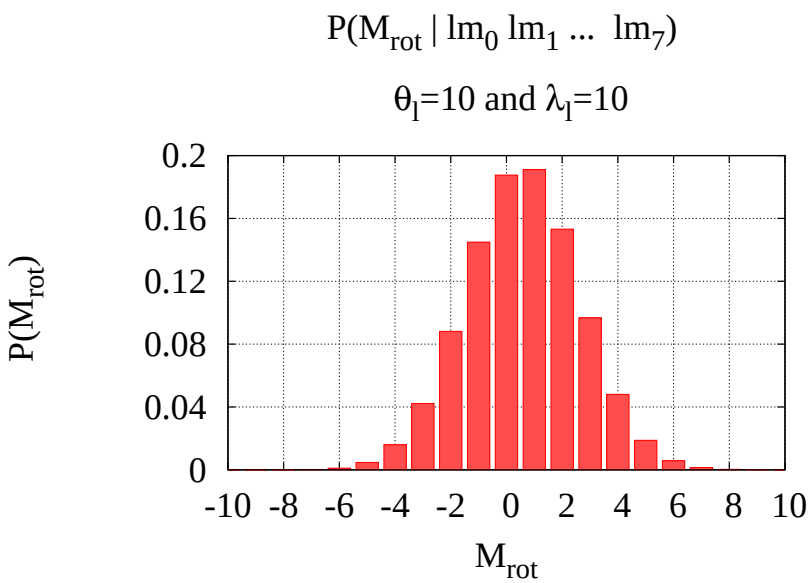


Figure 3.3: Rotation speed for phototaxy behaviour. The curve shows the probability distribution for variable M_{rot} when the sensor measures are those shown in Figure 3.2

3.2 Requirements Analysis

In this section, the Bayesian program proposed above is analysed, but it is not compared with the proposal in Chapter 1. As described above, this program deals with only one time step, and consequently it does not have the same functionality as the naive solution in the earlier chapter.

This analysis aims not only to emphasise the advantages of using state variables but also to introduce the procedure of space and time complexity analysis of Bayesian programs which will be used in subsequent chapters.

Computing the necessary memory space and time required for computation are connected. If a table is used to represent a probability distribution, it requires memory space to be stored, but access time is then very low. It is possible to reduce the required memory by describing a probability distribution by a function, where only its parameters are stored. On the other hand, this functional representation increases computation time. A table representation is more practical when it is desired to compare requirements of different filters.

When analysing the amount of required memory, we calculate roughly the space necessary to store all terms in the joint distribution as tables.

For time complexity analysis, we assume all distribution terms are normalised, and the normalisation calculation is not taken into account. Furthermore, all mathematical operations (like additions and multiplications) are considered to have similar costs.

Reduction of both required memory and time can be achieved by using particular representations and inference methods, but we try to analyse the general case, without

further assumptions. These conditions will be applied everywhere in this document, unless explicitly indicated otherwise.

3.2.1 Space Complexity

Keeping this program in memory means storing all the joint distribution terms in equation 3.2, as well as answers to questions. In addition to the parameters presented in Section 1.1, we introduce here the number of state variables N_s. The same number of cases n is assumed for all state variables.

The amount of memory required to store each element in the joint distribution is listed below. As indicated above, we consider all terms are stored in tables.

- The a priori term for state variables $P(S^j|\pi_s)$ contains n^{N_s} probability values.
- Each sensor model $P({}_kZ^j|S^j\pi_s)$ requires n^{N_s+1} probability values stored in memory, so for all N_z sensor models, $N_z n^{N_s+1}$ probability values are necessary.
- The motor command model $P(M^j|S^j\pi_s)$ is stored in memory using $n^{N_s+N_m}$ probability values.
- The result of the state estimation question in equation 3.3 is actually a probability distribution over states $P(S^j|\pi_s)$, so it contains n^{N_s} probability values.
- The answer to the motor command question as stated in equation 3.4 is a probability distribution over the motor variables $P(M^j|\pi_s)$, and it is stored using n^{N_m} probability values.

Summing all the above quantities, we have the expression for the whole amount of memory necessary to use the above Bayesian program:

$$\mathcal{M} = 2n^{N_s} + n^{N_m} + N_z n^{N_s+1} + n^{N_s+N_m}. \tag{3.10}$$

Without any assumptions regarding the number of variables, we can see that memory space has an upper bound of:

$$\mathcal{M} = O(N_z n^{N_s} + n^{N_s+N_m}). \tag{3.11}$$

The above expression clearly shows that the space required to store this Bayesian program increases according to the size of the sensor models (terms in blue) and the motor model (in brown). The space required increases linearly when increasing the number of observations, and exponentially with the number of state and motor variables.

Considering real applications, the number of observation variables is usually bigger than the number of state variables: the previous example illustrates this situation, with eight observation variables and only two state variables. In these cases, utilisation of sensor fusion schema, and consequent reduction for a linear dependence on space complexity, is especially useful.

3.2.2 Time Complexity

Two questions are asked of this joint distribution. To evaluate the time complexity of using this Bayesian program, it is necessary to know how many operations are required to calculate each question expression.

For the state estimation question, it is necessary to calculate equation 3.3, repeated below:

$$\begin{aligned} & P(S^j|_1 z^j{}_2 z^j \ldots_{N_z} z^j \pi_s) \\ & \quad \propto \quad P(S^j|\pi_s) \prod_{k=1}^{N_z} \left[P({}_k z^j | S^j \pi_s) \right] \quad . \end{aligned} \tag{3.12}$$

It can be seen that N_z multiplications are necessary for each value of each variable in the state space. The total number of multiplications is then $N_z n^{N_s}$.

Equation 3.5 shows that a summation over state space is necessary to solve the motor question:

$$\begin{aligned} & P(M^j|_1 z^j{}_2 z^j \ldots_{N_z} z^j \pi_s) \\ & \quad \propto \sum_{S^j} \left[\begin{array}{l} P(S^j|_1 z^j{}_2 z^j \ldots_{N_z} z^j \pi_s) \\ \times P(M^j|S^j \pi_s) \end{array} \right]. \end{aligned} \tag{3.13}$$

Using the state estimation calculated above, it is possible to reduce the motor question calculation to $n^{N_s} - 1$ sums plus n^{N_s} products, for each possible value of all motor variables. The total number of operations for the motor question is then $2n^{N_m+N_s} - n^{N_m}$.

To calculate both the estimation and motor questions, the number of necessary operations is given by:

$$\mathcal{C} = N_z n^{N_s} + 2n^{N_s+N_m} - n^{N_m}. \tag{3.14}$$

Without any further assumption in relation to the number of variables, the number of operations has an upper bound of:

$$\mathcal{C} = O(N_z n^{N_s} + n^{N_s+N_m}) \tag{3.15}$$

The above expression indicates that the time required grows linearly with the number of observation variables, and exponentially with the number of state and motor variables.

The number of observation variables in real applications is usually larger than the number of state variables, and this emphasises the importance of the linear dependence with N_z.

If sensor fusion is not applied, the number of observation variables is not present in expression of time complexity. But there is an exponential dependence on number of observation variables in space complexity. Application of sensor fusion adds a linear dependence in space complexity and reduces from exponential to linear dependence in

time complexity. The sensor fusion schema is actually a tradeoff between memory space and calculation effort, to be evaluated for each application.

3.2.3 Programming

As the state variables describe most relevant features of the specified robot task, it is expected that these variables are sufficient to define motor commands for situations involving execution of this specified task. Consequently, it is easier to relate motor commands to state variables than to observation variables.

Observations can be varied and dispersed, concerning several aspects of the robot and the environment, some of which are not relevant for the specified robot task. The sensor model allows us to concentrate on significant aspects of observations and their relation to state variables.

In comparison with the table proposed in Section 1.1, state variables simplify the activity of choosing motor commands for execution of the specified task according to environment situations.

Learning

Instead of writing all the terms in the joint distribution, it is possible to build them using learning techniques. In a Bayesian program, learning methods can be applied at several levels, from finding relevant variables to tuning parameters. In this particular program, parametric forms and their parameters can be learnt.

Simonin [2004] employed learning to build both sensor and motor models using the Bayesian Programming framework. The sensor model is first learnt off-line, and the motor model is then built, employing the sensor model: different motor commands are executed and changes in the state variables are evaluated. It is interesting to note that with this unique learning process, several simple behaviours were possible according to how the questions are asked.

Sensory-motor learning techniques in the context of a Bayesian robot programming framework have been studied by others [Lebeltel, 1999; Diard and Lebeltel, 2000; Raspail, 2002; Diard, 2003].

3.3 Discussion

3.3.1 Robotics

State variables and sensor fusion are used regularly in mobile robotics for several reasons. State variables are able to uncouple sensor and motor systems, thus allowing more sophisticated reasoning and consequent complex behaviours. Mapping of a robot's surroundings, for example, is achieved by means of state variables describing places or locations in this map.

Sensor fusion's main advantage is to increase the quality of state estimation. Redundancy in observations improves estimation precision and ensures robustness to failures.

When the sensors are complementary (for example, when each sensor provides information about one of several dimensions in the state space), the projection of the fusion result on each dimension is actually equivalent to each sensor observation. The fusion process itself does not add new information: if only one sensor measure is available, the result will depend on this sensor only.

Another useful characteristic is that a sensor observation model is easier to write in the form of a direct model, as is necessary for applying sensor fusion [Pradalier et al., 2003b].

3.3.2 Cognition

The cognitive interpretations of the mathematical hypotheses made in this chapter (as shown in decomposition equation 3.2) are the following:

- The Central Nervous System uses internal states to summarize the huge amount of sensory-motor information it shall process and to break the complexity of the required computation.

- When using sensor models, it is assumed that knowing the internal state, each sensory input may be approximately predicted independently of the other inputs:

$$\prod_{k=1}^{N_z} \left[P({}_kZ^j | S^j \pi_s) \right].$$

- Another assumption is that, knowing the internal states, the adequate motor commands may be predicted:

$$P(M^j | S^j \pi_s).$$

The first hypothesis is hardly challenged nowadays. But we would like to remark that, as the world exists in its whole complexity, nothing proves that this complexity arises from less complex states. From our point of view, the observer or modeller has its own internal states for the world and what is usually called *states of the world* is the verbalised or formalised version of these internal states.

It is a presumption to suppose that the modeller's internal states have any thing in common with the internal states of the observed being, even if they both get sensory inputs from the same environment, and they may solve alike problems. Does an entomologist share part of its representation of the world with the ant he is observing?

We understand that this confusion is at the origin of numerous misinterpretations. For instance, when analysing navigation strategies of animals, we often project geometrical interpretations that may be rather the result of our own mathematical point of view of the problem.

We believe that the essence of internal states is in the second hypothesis. The internal states explain the world in the sense that their exact values make possible to predict the sensory inputs independently. Internal states are then the cause explaining the sensory inputs (the consequences): knowing the "cause", the "consequences" are independent. How the Central Nervous System would discern these explanatory variables is still an open question, not addressed here but subject of an ongoing thesis in our group.

In sections 3.2.1 and 3.2.2 a practical interest of this second hypothesis is presented. Instead of exponential growth according to the number of sensory variables in both space and time complexities, we have an exponential growth with the number of internal state variables. The number of possible internal states is much smaller than the number of possible sensory situations, which make this reduction one of the major simplifications towards having a tractable problem.

This simplification mechanism may be what is shared by the ant and the entomologist. Both have internal states to explain their sensory inputs and to reduce the demands on memory and processing power. However, they certainly do not have the same sensory input and, as a consequence, they would most probably not share the same internal states.

The third hypothesis is nearly as important for simplification than the second one, but we prefer to postpone its discussion: it will be precised and refined in chapters 5, 6 and 7.

3.4 Summary and Open Questions

We have shown in this chapter that state variables describe relevant aspects of robot surroundings and of the specified task, and they reduce the difficulty of programming the robot. This simplification is possible because of the uncoupling of observations and motor variables.

Actually, state variables are related to observations by means of the sensor model, which is defined based on sensor-supplied information:

$$P(Z^j|S^j\pi_s).$$

Conditional independence between sensors when the phenomenon being measured is known leads to the sensor fusion equation:

$$P(Z^j|S^j\pi_s) = \prod_{k=1}^{N_z} \Big[P({}_kZ^j|S^j\pi_s)\Big].$$

In addition, redundant and complementary information coming from sensor observations can be combined into state variables by applying sensor fusion.

Furthermore, state variables are related to motor commands. The motor model is employed to express these relations:

$$P(M^j|S^j\pi_s).$$

The aspects shown above were described and analysed for a single time step, meaning that dependence with time is not solved by the addition of state variables. This is the subject of the next chapter.

Chapter 4

Bayes Filter

Key Points:

- **Markov assumption and stationarity for reducing time dependence**
- **Partially uncouple dependencies between time instants without forgetting the past.**

The solution to the ultimate question proposed in Section 1.1 presented a major problem related to time dependence. The amount of memory necessary to store the table increases exponentially with time and with the number of variables.

It is proposed in this chapter to apply a Bayes filter structure to uncouple dependencies partially between time instants. It is not desired to lose all information about the past but to limit time dependence to a reasonable depth. This limit is imposed here by applying the Markov assumption. Dependence in time is also reduced by assuming models are stationary in time.

Information about the past is summarised, and present-time considerations are employed to estimate current variable values. This is achieved by a recursive calculation that reduces computational effort considerably.

The Bayes filter proposed in this chapter includes the motor command model in the recursive filter calculation. This allows the filter to provide an answer to the ultimate question, and it represents our first proposal for a solution.

4.1 Bayesian Program

In this chapter, we consider again variables with their history in time. We saw in Chapter 1 that a bigger table expressing dependencies between variables was necessary at each new time step.

We aim here to define the dependence between variables at a given time and to keep these dependencies stationary: they do not change with time. To maintain a relationship with the past, the variables considered include present- and past-time variables.

The dependencies between variables employed in the previous chapter are used here with some differences: the sensor and motor models are still used, but a dynamic model for the state variables is added. Whenever useful, past-time variables are included.

Utilisation is also very similar, with state estimation and motor command questions. To reduce calculation, a recursive loop is used to answer these questions.

4.1.1 Description

Additional Relevant Variables

As in the previous chapter, the variables involved in this program are the set of observation variables Z, the set of motor command variables M and the set of state variables S. However, in contrast to the previous chapter, all time instants, from zero to an arbitrary time instant t, are considered. The whole set of pertinent variables is therefore $M^{0:t}$, $Z^{0:t}$ and $S^{0:t}$.

Decomposition

The Bayes filter joint distribution involves all relevant variables:

$$\begin{aligned} & P(M^{0:t} S^{0:t} Z^{0:t} | \pi_f) \\ & = \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S^j | S^{j-1} M^{j-1} \pi_f) \\ \times P(Z^j | S^j \pi_f) \\ \times P(M^j | S^j M^{j-1} \pi_f) \end{array} \right] \\ \times P(M^0 S^0 Z^0 | \pi_f). \end{array} \right. \end{aligned} \tag{4.1}$$

The first term inside the product is called the dynamic model, and it expresses our knowledge about how the state variables evolve in time. The state variables at time instant j depend on the state variables at the previous time step S^{j-1} and on how the robot moved in the previous time step M^{j-1}. As the state variables describe world features relevant to the robot's task, the dynamic model is responsible for identifying eventual changes in world features as time passes and according to the robot's action over the environment.

The second term inside the product is the sensor model, already detailed in the previous chapter.

The third term inside the product is the motor model, but it is a different pattern of motor model from that presented in the previous chapter. Actually, the motor commands at time instant j depend not only on the state variables at time instant j, as described earlier, but also on the motor commands at the previous time instant M^{j-1}.

Outside the product, there is a term for the initial conditions, i.e. the distribution over relevant variables at time instant $j = 0$.

We can see that the first-order Markov assumption is present in both the dynamic and motor models: time dependence has a depth of one time step. The product of terms

is a consequence of the stationarity assumption: models do not change with time. They are defined over different variables at each time instant, but the relationships between these variables remain the same for all time steps.

4.1.2 Utilisation

The Bayes filter described in the above joint distribution allows recursive utilisation. Mathematical proof of this property of recursiveness is shown below, where the equations effectively used in the recursive loop are described in detail. The proof is an extended version of the one presented by Coué [2003].

Recursive calculation in the above Bayes filter consists of asking three questions, sequentially, to filter the joint distribution. The first question is called *prediction*, and in this part of the calculation loop, the information about the past is updated using the dynamic model. It therefore consists of a probabilistic question with states to be estimated at the present time (S^t) on the left side and variables known up to now on the right side: all past-time observation values and motor commands, $z^{0:t-1}$ and $m^{0:t-1}$, respectively.

$$P(S^t|m^{0:t-1}z^{0:t-1}\pi_f).$$

The second question is called *estimation*, and it aims to update the prediction question result with evidence given by observations. In the estimation question, we have on the left side the present state variables to be estimated S^t, and on the right side, the last observation z^t is added to the information already available, so we have $z^{0:t}$ and $m^{0:t-1}$.

$$P(S^t|m^{0:t-1}z^{0:t}\pi_f).$$

Once state variable estimations are computed, it is possible to apply this estimation for deciding which motor commands are most relevant for the present estimated situation. The *motor command* question has on the left side the motor variables at the present time M^t and on the right side the same set of variables used in estimation, namely $z^{0:t}$ and $m^{0:t-1}$.

$$P(M^t|m^{0:t-1}z^{0:t}\pi_f).$$

Developing the prediction question to obtain the expression to solve it, we apply the inference rule (as shown in equation **??**) to find:

$$\begin{aligned} &P(S^t|m^{0:t-1}z^{0:t-1}\pi_f) \\ &\quad\propto \sum_{\begin{bmatrix} S^{0:t-1} \\ M^t \\ Z^t \end{bmatrix}} \begin{bmatrix} \prod_{j=1}^{t} \begin{bmatrix} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^jM^{j-1}\pi_f) \end{bmatrix} \\ \times P(M^0S^0Z^0|\pi_f) \end{bmatrix} . \end{aligned} \tag{4.2}$$

To simplify this expression, the terms at time step t are taken out of the time product, and, applying the normalisation rule, they sum to one, as shown in the following equations.

$$
\begin{aligned}
&P(S^t|m^{0:t-1}z^{0:t-1}\pi_f) \\
&\quad\propto \sum_{\begin{bmatrix} S^{0:t-1} \\ M^t \\ Z^t \end{bmatrix}} \begin{bmatrix} P(S^t|S^{t-1}m^{t-1}\pi_f) \\ \times P(Z^t|S^t\pi_f) \times P(M^t|S^t m^{t-1}\pi_f) \\ \prod_{j=1}^{t-1} \begin{bmatrix} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^j M^{j-1}\pi_f) \end{bmatrix} \\ \times P(M^0 S^0 Z^0|\pi_f) \end{bmatrix} \\
&\quad\propto \sum_{S^{0:t-1}} \begin{bmatrix} \overbrace{\sum_{Z^t}\{P(Z^t|S^t\pi_f)\}}^{=1} \times \overbrace{\sum_{M^t}\{P(M^t|S^t m^{t-1}\pi_f)\}}^{=1} \\ \times P(S^t|S^{t-1}m^{t-1}\pi_f) \\ \prod_{j=1}^{t-1} \begin{bmatrix} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^j M^{j-1}\pi_f) \end{bmatrix} \\ \times P(M^0 S^0 Z^0|\pi_f) \end{bmatrix} \\
&\quad\propto \sum_{S^{0:t-1}} \begin{bmatrix} P(S^t|S^{t-1}m^{t-1}\pi_f) \\ \prod_{j=1}^{t-1} \begin{bmatrix} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^j M^{j-1}\pi_f) \end{bmatrix} \\ \times P(M^0 S^0 Z^0|\pi_f) \end{bmatrix}
\end{aligned}
\tag{4.3}
$$

In this last expression, we can see that all past evolution of state variables is included in the prediction question answer by means of the product of terms from $j = 1$ to $j = t - 1$. Information for the current time instant is present by means of the dynamic model in the first line of the expression.

To calculate the estimation expression, we apply the general inference rule of equation **??** to obtain:

$$
\begin{aligned}
&P(S^t|z^{0:t}m^{0:t-1}\pi_f) \\
&\quad\propto \sum_{\begin{bmatrix} S^{0:t-1} \\ M^t \end{bmatrix}} \begin{bmatrix} \prod_{j=1}^{t} \begin{bmatrix} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^j M^{j-1}\pi_f) \end{bmatrix} \\ \times P(M^0 S^0 Z^0|\pi_f) \end{bmatrix} .
\end{aligned}
\tag{4.4}
$$

Normalisation can also help to simplify this expression. First, we take out from the time product the terms of time step t. Then we simplify for variable M^t, as shown below.

$$
\begin{aligned}
&P(S^t|z^{0:t}m^{0:t-1}\pi_f) \\
&\propto \sum_{\left[\begin{array}{c} S^{0:t-1} \\ M^t \end{array}\right]} \left[\begin{array}{l} P(S^t|S^{t-1}m^{t-1}\pi_f) \\ \times P(z^t|S^t\pi_f) \times P(M^t|S^t m^{t-1}\pi_f) \\ \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^j M^{j-1}\pi_f) \end{array}\right] \\ \times P(M^0 S^0 Z^0|\pi_f) \end{array}\right] \\
&\propto \sum_{\left[\begin{array}{c} S^{0:t-1} \end{array}\right]} \left[\begin{array}{l} \overbrace{\sum_{M^t} \{P(M^t|S^t m^{t-1}\pi_f)\}}^{=1} \\ \times P(S^t|S^{t-1}m^{t-1}\pi_f) \times P(z^t|S^t\pi_f) \\ \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^j M^{j-1}\pi_f) \end{array}\right] \\ \times P(M^0 S^0 Z^0|\pi_f) \end{array}\right] \\
&\propto \left| \begin{array}{l} P(z^t|S^t\pi_f) \\ \times \sum_{\left[\begin{array}{c} S^{0:t-1} \end{array}\right]} \left[\begin{array}{l} P(S^t|S^{t-1}m^{t-1}\pi_f) \\ \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^j M^{j-1}\pi_f) \end{array}\right] \\ \times P(M^0 S^0 Z^0|\pi_f) \end{array}\right] \end{array} \right. .
\end{aligned}
\tag{4.5}
$$

In the last expression above, we can see that estimation is a function of the prediction expression stated in equation 4.3. We have then:

$$
\begin{aligned}
&P(S^t|z^{0:t}m^{0:t-1}\pi_f) \\
&\propto \left| \begin{array}{l} P(z^t|S^t\pi_f) \\ \times P(S^t|m^{0:t-1}z^{0:t-1}\pi_f) \end{array} \right. .
\end{aligned}
\tag{4.6}
$$

As stated earlier, the estimation question (in blue) updates the prediction with available observations, as can be seen clearly in the expression above. The observation model in the first line corrects the prediction expression, in green.

So that the recursive calculation can loop, it is still necessary to show that by using the estimation expression of the previous time step, it is possible to calculate the prediction. Before this, let us see how the motor commands for time instant j are calculated.

For an expression answering the motor question, we apply the inference rule in equation **??** and we find:

$$P(M^t|z^{0:t}m^{0:t-1}\pi_f)$$

$$\propto \sum_{S^{0:t}} \left[\begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^jM^{j-1}\pi_f) \end{array} \right] \\ \times P(M^0S^0Z^0|\pi_f) \end{array} \right]. \tag{4.7}$$

Normalisation does not help to simplify this expression, but it is possible to rearrange the terms so that it becomes a function of the estimation expression. First, the terms for time step t are taken out of the time product, then the estimation expression is substituted.

$$\begin{aligned} & P(M^t|z^{0:t}m^{0:t-1}\pi_f) \\ & \propto \sum_{S^{0:t}} \left[\begin{array}{l} P(S^t|S^{t-1}m^{t-1}\pi_f) \\ \times P(z^t|S^t\pi_f) \times P(M^t|S^tm^{t-1}\pi_f) \\ \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^jM^{j-1}\pi_f) \end{array} \right] \\ \times P(M^0S^0Z^0|\pi_f) \end{array} \right] \\ & \propto \sum_{S^t} \left[\begin{array}{l} P(M^t|S^tm^{t-1}\pi_f) \\ \times P(z^t|S^t\pi_f) \\ \times \sum_{S^{0:t-1}} \left[\begin{array}{l} P(S^t|S^{t-1}m^{t-1}\pi_f) \\ \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^jM^{j-1}\pi_f) \end{array} \right] \\ \times P(M^0S^0Z^0|\pi_f) \end{array} \right] \end{array} \right] \\ & \propto \sum_{S^t} \left[\begin{array}{l} P(M^t|S^tm^{t-1}\pi_f) \\ \times P(S^t|m^{0:t-1}z^{0:t}\pi_f) \end{array} \right]. \end{aligned} \tag{4.8}$$

It can be seen that a summation over state variables is necessary for calculating the motor command expression (in brown). Actually, the state variables have no assigned values, and the motor model is weighted by the estimated distribution (in blue) for all possible values of all state variables. From the resulting probability distribution, values for the set of motor commands are drawn, and m^t becomes known.

This ends all calculations for the present time instant and the next calculation loop begins with prediction question. We want to show that prediction is calculated using the estimation and motor command values from the previous time instant.

The following equations develop the prediction question expression in equation 4.3. First, the terms at time step $(t-1)$ are taken out of the time product, then the expression of estimation as in equation 4.6 is replaced by the estimation question.

$$P(S^t|m^{0:t-1}z^{0:t-1}\pi_f)$$

$$
\begin{aligned}
&\propto \sum_{S^{0:t-1}} \left[\begin{array}{l} P(S^t|S^{t-1}m^{t-1}\pi_f) \\ \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^jM^{j-1}\pi_f) \end{array} \right] \\ \times P(M^0S^0Z^0|\pi_f) \end{array} \right] \\
&\propto \sum_{S^{0:t-1}} \left[\begin{array}{l} P(S^t|S^{t-1}m^{t-1}\pi_f) \\ \times P(S^{t-1}|S^{t-2}m^{t-2}\pi_f) \\ \times P(z^{t-1}|S^{t-1}\pi_f) \times P(m^{t-1}|S^{t-1}m^{t-2}\pi_f) \\ \prod_{j=1}^{t-2} \left[\begin{array}{l} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^jM^{j-1}\pi_f) \end{array} \right] \\ \times P(M^0S^0Z^0|\pi_f) \end{array} \right] \\
&\propto \sum_{S^{t-1}} \left[\begin{array}{l} P(S^t|S^{t-1}m^{t-1}\pi_f) \\ \times P(m^{t-1}|S^{t-1}m^{t-2}\pi_f) \times \\ \overbrace{P(z^{t-1}|S^{t-1}\pi_f)}^{\textit{Estimation at time step } t-1} \\ \times \sum_{S^{0:t-2}} \left[\begin{array}{l} P(S^{t-1}|S^{t-2}m^{t-2}\pi_f) \\ \prod_{j=1}^{t-2} \left[\begin{array}{l} P(S^j|S^{j-1}M^{j-1}\pi_f) \\ \times P(Z^j|S^j\pi_f) \\ \times P(M^j|S^jM^{j-1}\pi_f) \end{array} \right] \\ \times P(M^0S^0Z^0|\pi_f) \end{array} \right] \end{array} \right] \\
&\propto \sum_{S^{t-1}} \left[\begin{array}{l} P(S^t|S^{t-1}m^{t-1}\pi_f) \\ \times P(m^{t-1}|S^{t-1}m^{t-2}\pi_f) \\ \times P(S^{t-1}|z^{0:t-1}m^{0:t-2}\pi_f) \end{array} \right]. \qquad (4.9)
\end{aligned}
$$

This last expression shows interesting characteristics. Past-time estimation (in blue) is equivalent to a summary of all past information about states. It is updated with the motor command model in the second line, with values drawn from the motor question results.

In the first line, the dynamic model establishes the relation between states at past and present time instants. The summation over previous-time state variables reflects the fact that state values are never known: state estimation at time instant $t-1$ (updated by the motor models) weights the dynamic model to predict the state at the present time.

The sequence of questions is better visualised in the diagram shown in Figure 4.1. The dotted line marks the transition from time step $t-1$ to time step t. Prediction (green square) is the first question asked at each time step, and it employs state information from the previous time step, as well as motor variable values.

The estimation question (blue square) is then asked to the joint distribution of the filter: observation values z^t are taken into account to update state prediction. Next, the motor question (brown square) is asked, and from the answer to that question, the values of motor commands m^t are decided. Correspondence between equations and diagram can be easily made by associating colours. This correspondence will be applied

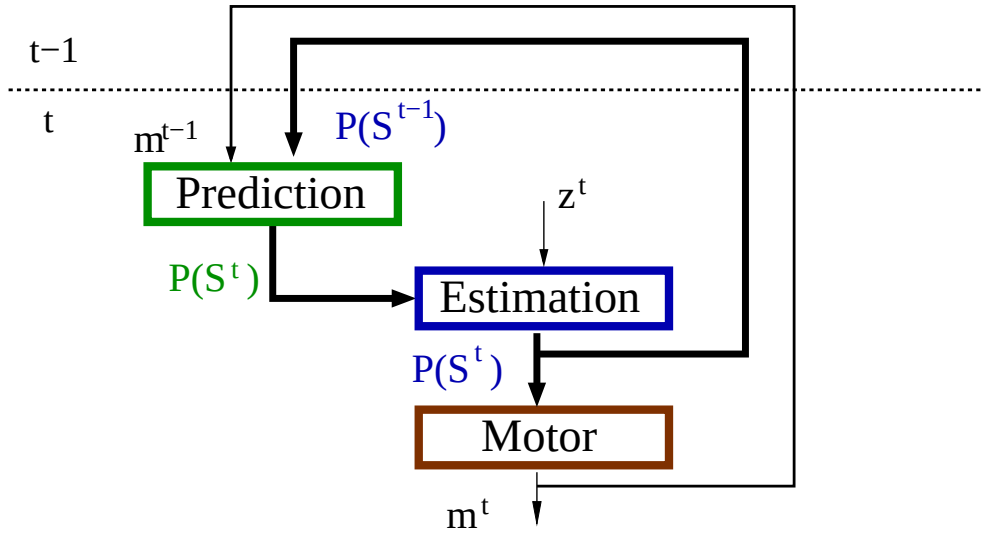


Figure 4.1: Diagram of Bayes filter utilisation. The dotted line indicates the transition from time instant $t-1$ to time instant t. Expressions for prediction, estimation and motor questions can be found in the text.

as often as possible in the following chapters, where this type of diagram will be used to illustrate utilisation.

4.1.3 Illustration Examples

In this chapter, we introduce the simulated environment applied as illustration here and in the following chapters. Initially, a simplified version is presented in Example 1 to demonstrate how a simple Bayes filter works. Then, the improved version used in following chapters is described in Example 2, along with the results of the Bayes filter described in Section 4.1.

Example 1

In this first example, our robot is inside an arena, without obstacles. It can move freely, although it cannot go through the walls. Our robot is not alone: a bad predator is there and its goal is to capture the robot. Robot behaviour in this situation is very simple and consists of escaping from the predator. Figure 4.2 illustrates robot and predator.

Bayesian Program

Robot movements depend basically on the predator's position. The relevant state variables are therefore predator direction and distance in relation to the robot, X_{dir} and X_{dist}, respectively. Distance is an integer number from 0 (very close) to 4 (far away). The direction variable value assumes one of seven values: -180, -120, -60, 0, 60, 120, 170. Each value is measured in relation to the robot's forward direction. If the predator

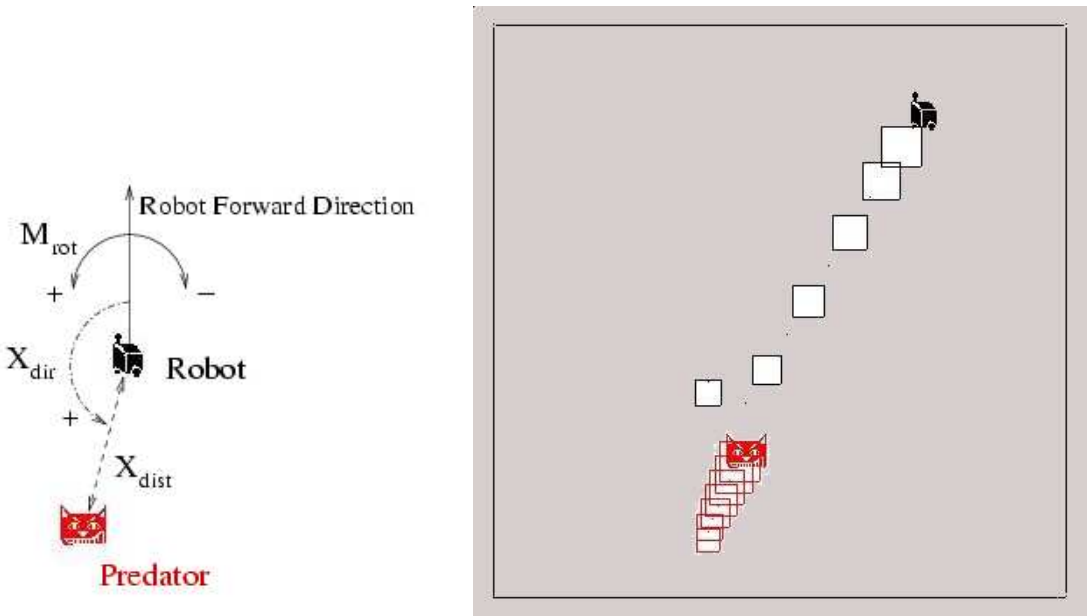


Figure 4.2: Simulation of Example 1: robot (black) and predator (red). At left, variables describing the predator position and rotation speed signal convention are also shown. At right, the robot escaping from the predator.

direction is equal to zero, the predator is in front of the robot; When the predator direction is positive, the predator is on the left side of the robot, and when the direction has a negative value, the predator is on the robot's right side. When predator direction is equal to –180 or 170, the predator is behind the robot. A graphical description of these variables is shown in the left part of Figure 4.2.

In a simulated environment, real sensors are not necessary, and observation variables can be freely chosen. However, we try to select observation variables that are as nearly as possible those of a real situation. In this example, the observation variables are the observed direction and distance of the predator (ZX_{dir} and ZX_{dist}, respectively). It is reasonable to assume that they could be derived by processing data from laser range finders or vision cameras.

The motor control of the robot in this simulated environment is chosen to be similar to a real robot. Wheeled robots are often controlled by the wheel speeds, but it is also possible to command it using translation and rotation speed, as shown in equation 2.9. As these variables are usually more convenient, motor variables in this example are M_{rot} and M_{trans}, for rotation and translation speed, respectively. Translation speed is an integer number from 0 (robot does not advance) and 4 (robot advances at full speed). Rotation speed can be assigned one of seven values: -180, -120, -60, 0, 60, 120, 180 degrees per time step. Rotation speed equal to 0 means no rotation; a positive value means turning left and a negative value means turning right.

Following the general joint distribution shown in equation 4.1, for this example we have:

$$
\begin{aligned}
&P(X_{dist}^{0:t} X_{dir}^{0:t} ZX_{dist}^{0:t} ZX_{dir}^{0:t} M_{rot}^{0:t} M_{trans}^{0:t} | \pi_{E1}) \\
&= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(X_{dist}^{j} | X_{dist}^{j-1} \pi_{E1}) \times P(X_{dir}^{j} | X_{dir}^{j-1} M_{rot}^{j-1} \pi_{E1}) \\ \times P(ZX_{dist}^{j} | X_{dist}^{j} \pi_{E1}) \times P(ZX_{dir}^{j} | X_{dir}^{j} \pi_{E1}) \\ \times P(M_{rot}^{j} | X_{dist}^{j} X_{dir}^{j} \pi_{E1}) \\ \times P(M_{trans}^{j} | X_{dist}^{j} \pi_{E1}) \end{array} \right] \\ \times P(X_{dist}^{0} X_{dir}^{0} ZX_{dist}^{0} ZX_{dir}^{0} M_{rot}^{0} M_{trans}^{0} | \pi_{E1}) \end{array} \right| .
\end{aligned} \tag{4.10}
$$

The terms in the first line of this distribution are the dynamic models, and they show interesting properties. The predator distance variable, for example, depends only on the distance at the previous time. This property illustrates how preliminary knowledge about the predator can be expressed: the predator aims to follow the robot and consequently, its distance tends to decrease with time . The dynamic model consists of a probability table built with the non-normalised function below. Normalisation takes place after the table is filled.

$$
P(X_{dist}^{j} | X_{dist}^{j-1} \pi_{E1}) = \begin{cases} 10, \text{if} X_{dist}^{j} = X_{dist}^{j-1} - 1; \\ 10, \text{if} X_{dist}^{j} = X_{dist}^{j-1}; \\ 1, \text{otherwise.} \end{cases} \tag{4.11}
$$

The predator direction, on the other hand, depends on both its direction and the robot's rotation speed at the previous time. This illustrates another utility of the dynamic model: information about the environment is updated in relation to the robot's actions in the environment. The parametric form for this term is then chosen to be a normal curve, with a mean value that is the correction of the direction at the previous time instant according to the robot's rotation speed. The standard deviation is then responsible for expressing uncertainty about this value, and here it is considered constant.

$$
P(X_{dir}^{j} | X_{dir}^{j-1} M_{rot}^{j-1} \pi_{E1}) = \mathcal{G}(X_{dir}^{j}), \mu = X_{dir}^{j-1} - M_{rot}^{j-1}, \sigma_{dyn}. \tag{4.12}
$$

The sensor models are very simple in this example, as both state and observation variables have the same type of information. With a real robot, imprecision and uncertainty according to the method employed to process sensor raw data can be modelled. In this example, a normal curve is employed as the sensor model: the mean value is equal to the state variable, and the standard deviation is a constant value. For predator distance we have:

$$
P(ZX_{dist}^{j} | X_{dist}^{j} \pi_{E1}) = \mathcal{G}(ZX_{dist}^{j}), \mu = X_{dist}^{j}, \sigma_{z}. \tag{4.13}
$$

The motor model determines the robot's actions based on the state variables. The robot's rotation speed depends on the predator direction and distance: for example, if the predator is near and to the right, the robot must turn left.

The rotation speed model is chosen to be a normal curve: its mean is a function of the distance and direction of the predator, and its standard deviation is considered

constant. Reaction to the predator's presence begins when the distance is three or less, and the robot is at the right if the direction is negative. The mean of the normal curve then changes accordingly: if the distance is three or less and the direction is positive or zero, the mean is equal to the opposite direction turning left; if the distance is three or less and the direction is negative, the mean is equal to the opposite direction turning right; if the distance is greater than three, there is no change in the robot's direction (mean equal to zero).

$$\begin{aligned} &P(M^j_{rot}|X^j_{dist}X^j_{dir}\pi_{E1}) = \mathcal{G}(M^j_{rot}), \mu = f(X^j_{dist}, X^j_{dir}), \sigma_{rot}, \\ &f(X^j_{dist}, X^j_{dir}) = \begin{cases} X^j_{dir} - 180, \text{if } X^j_{dist} < 3 \text{ and } X^j_{dir} \geq 0; \\ X^j_{dir} + 180, \text{if } X^j_{dist} < 3 \text{ and } X^j_{dir} < 0; \\ 0, \text{otherwise.} \end{cases} \end{aligned} \tag{4.14}$$

The translation speed is simpler: if the predator is near, go quickly; if it is distant, go more slowly. Its parametric form is also a normal curve.

The first step in utilisation is to calculate the prediction question:

$$P(X^{0:t}_{dist}X^{0:t}_{dir}|zx^{0:t-1}_{dist}zx^{0:t-1}_{dir}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\pi_{E1}).$$

Then we ask the estimation question:

$$P(X^{0:t}_{dist}X^{0:t}_{dir}|zx^{0:t}_{dist}zx^{0:t}_{dir}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\pi_{E1}).$$

Finally, the motor question is calculated:

$$P(M^t_{rot}M^t_{trans}|zx^{0:t}_{dist}zx^{0:t}_{dir}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\pi_{E1}).$$

Simulations Results

The right part of Figure 4.2 shows the robot escaping from the predator. Squares of different sizes are employed to show the path followed by robot and predator. The centre of the square shows the position and square size is related to time: a big square shows a recent position and a small square indicates a position in the past. Squares have the same outline colour as the corresponding object (black for the robot, and red for the predator).

The initial position of the robot is indicated by the smallest black square, and the initial position of predator corresponds to the smallest red square. In the right part of Figure 4.2, the initial condition shows the predator near the robot. The robot's initial heading direction is facing the predator for this simulation run.

At time instant $t = 1$, the prediction question is asked of the joint distribution, but there is no estimation available from the previous time step. Actually, the initial condition terms in the joint distribution are all information regarding time instant $t = 0$. The prediction curve in Figure 4.3 results from applying the dynamic model to the initial conditions.

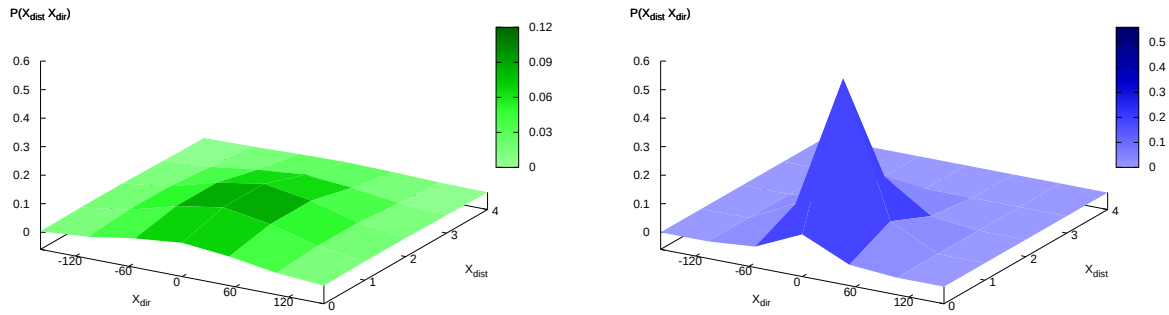


Figure 4.3: States Prediction (left, in green) and Estimation (right, in blue) for time instant $t = 1$, simulation for Example 1, shown in right part of figure 4.2.

It can be seen that in the prediction curve, the predator is situated facing the robot (values of direction near to zero are more probable) and near the robot (small values of distance variable have higher probability values). However, there is not much certainty about this information, as the prediction curve is flat and spread out, meaning that there is much uncertainty regarding the position of the predator.

The estimation curve at time instant $t = 1$ applies the observation available about the predator position (it is approaching the robot, chasing it) to update the prediction. As can be seen in the estimation curve in Figure 4.3, the direction and distance of the predator becomes more certain. The certainty in estimation depends mostly on how accurate the observation of the sensor model is considered to be.

Based on the estimation result, an appropriate motor command distribution is calculated to solve the motor question. In Figure 4.4, the motor command distribution curves are shown. The rotation speed is on the left side: the robot must rotate to its left, so that it does not face the predator. The translation speed curve on the right indicates that the robot must move quickly, as the predator is close. From these probability distributions, motor command values are drawn ($m^1_{rot} = 90$ and $m^1_{trans} = 4$), and these values together with state estimations are employed at the next time step to calculate the next prediction.

Example 2

In this second example, our robot is in the same arena, but now a prey and a nest are also present. When in the nest, also called home, the robot is safe from the predator.

Robot behaviour now is more complex, and it consists of escaping from the predator, chasing the prey and eventually going home. Figure 4.6(a) illustrates this arena.

Bayesian Program

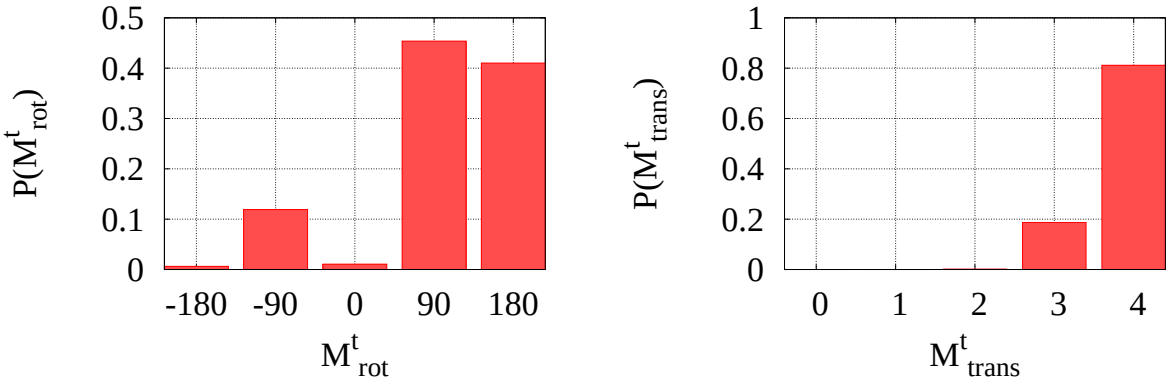


Figure 4.4: Motor commands distribution for time instant t=1, simulation for Example 1, shown in right part of figure 4.2.

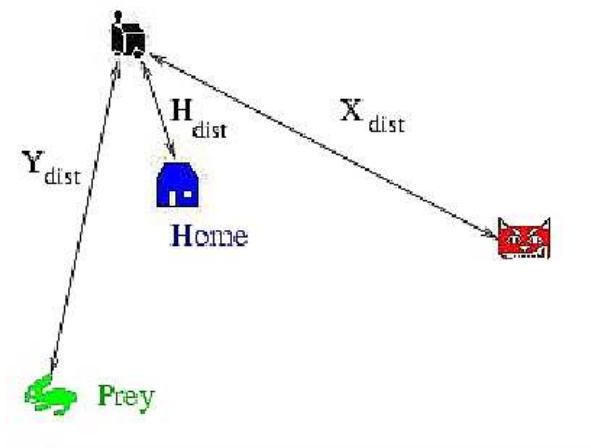


Figure 4.5: Illustration of simulation arena for Example 2, showing robot (black), predator (red), prey (green) and home (blue).

State variables describe the world features most relevant to the robot's specified behaviour. Following the reasoning in the first example, they are the direction and distance of predator, prey and home. Figure 4.5 shows the distance-related variables graphically, as direction variables are measured as in the first example. For the sake of brevity in the equations, the set of state variables will be denoted as S_{E2}:

$$S^j_{E2} = \{H^j_{dir} H^j_{dist} X^j_{dist} X^j_{dir} Y^j_{dist} Y^j_{dir}\}.$$

Observation variables are of the same type as state variables. This choice allows us to simplify the sensor model, and it is reasonable from a practical point of view. The set of observation variables for this example is denoted as Z_{E2}:

$$Z^j_{E2} = \{ZH^j_{dir} ZX^j_{dist} ZY^j_{dist} ZH^j_{dist} ZX^j_{dir} ZY^j_{dist}\}.$$

The motor command variables are the same as in the previous example, rotation speed M_{rot} and translation speed M_{trans}.

For all variables, the ranges and the number of cases are the same as for the first example.

The joint distribution for this example is then:

$$
\begin{aligned}
&P(S_{E2}^{0:t} Z_{E2}^{0:t} M_{rot}^{0:t} M_{trans}^{0:t} | \pi_{E2}) \\
&= \left| \begin{array}{l} \Pi_{j=1}^{t} \left[\begin{array}{l} P(H_{dist}^{j} | H_{dist}^{j-1} \pi_{E2}) \times P(H_{dir}^{j} | H_{dir}^{j-1} M_{rot}^{j-1} \pi_{E2}) \\ \times P(X_{dist}^{j} | X_{dist}^{j-1} \pi_{E2}) \times P(X_{dir}^{j} | X_{dir}^{j-1} M_{rot}^{j-1} \pi_{E2}) \\ \times P(Y_{dist}^{j} | Y_{dist}^{j-1} \pi_{E2}) \times P(Y_{dir}^{j} | Y_{dir}^{j-1} M_{rot}^{j-1} \pi_{E2}) \\ \times P(ZH_{dir}^{j} | H_{dir}^{j} \pi_{E2}) \times P(ZH_{dist}^{j} | H_{dist}^{j} \pi_{E2}) \\ \times P(ZX_{dist}^{j} | X_{dist}^{j} \pi_{E2}) \times P(ZX_{dir}^{j} | X_{dir}^{j} \pi_{E2}) \\ \times P(ZY_{dist}^{j} | Y_{dist}^{j} \pi_{E2}) \times P(ZY_{dist}^{j} | Y_{dist}^{j} \pi_{E2}) \\ \times P(M_{rot}^{j} | Y_{dist}^{j} Y_{dir}^{j} X_{dist}^{j} X_{dir}^{j} H_{dist}^{j} H_{dir}^{j} \pi_{E2}) \\ \times P(M_{trans}^{j} | Y_{dist}^{j} X_{dist}^{j} H_{dist}^{j} \pi_{E2}) \end{array} \right] \\ \times P(S_{E2}^{0} Z_{E2}^{0} M_{rot}^{0} M_{trans}^{0} | \pi_{E2}). \end{array} \right.
\end{aligned}
\tag{4.15}
$$

Inside the product, we can see that the first line deals with the dynamic models for home-related variables. The home is considered fixed in space, so the distance dynamic model is responsible for increasing the uncertainty in relation to distance. This is obtained by a normal function; its mean is the distance at the previous time, and its standard deviation is fixed.

$$P(H_{dist}^{t} | H_{dist}^{t-1} \pi_{E2}) = \mathcal{G}(H_{dist}^{t}), \mu = H_{dist}^{t-1}, \sigma_{dyn}. \tag{4.16}$$

For the home direction dynamics, a correction in relation to the robot's displacement is employed, as described in Example 1 for the predator direction dynamic model.

$$P(H_{dir}^{t} | H_{dir}^{t-1} M_{rot}^{t-1} \pi_{E2}) = \mathcal{G}(H_{dir}^{t}), \mu = H_{dir}^{t-1} - M_{rot}^{t-1}, \sigma_{dyn}. \tag{4.17}$$

The second line shows the dynamic model for the predator-related state variables. The same models described in Example 1 are used here.

In the next line, the dynamic models for the prey-related state variables include the available knowledge about the prey displacement: when pursued, a prey tries to escape, so the distance between the robot and the prey can increase with time. However, if the robot is chasing it, the distance can also decrease. To model this situation in a simpler way, the non-normalised function below is employed to fill a probability table. Normalisation is executed afterwards.

$$P(Y_{dist}^{t} | Y_{dist}^{t-1} \pi_{E2}) = \left\{ \begin{array}{l} 10, \text{if} Y_{dist}^{j} = Y_{dist}^{j-1} - 1; \\ 10, \text{if} Y_{dist}^{j} = Y_{dist}^{j-1}; \\ 10, \text{if} Y_{dist}^{j} = Y_{dist}^{j-1} + 1; \\ 1, \text{otherwise}. \end{array} \right. \tag{4.18}$$

No assumption regarding prey direction when escaping was included in the dynamic model for the direction state variable. Similarly to the home and predator direction dynamics, only a correction in relation to the previous robot rotation is considered:

$$P(Y^t_{dir}|Y^{t-1}_{dir}M^{t-1}_{rot}\pi_{E2}) = \mathcal{G}(Y^t_{dir}), \mu = Y^{t-1}_{dir} - M^{t-1}_{rot}, \sigma_{dyn}. \tag{4.19}$$

Following the joint distribution, the next line shows the observation models, initially for home-related variables, then for predator and prey. As in Example 1, sensor and state variables have similar types, and there is no transformation in the sensor model. They are all defined as normal; their means are the state variables and their standard deviations indicate the imprecision and uncertainty associated with sensor measures and data processing.

$$P(ZH^j_{dist}|H^j_{dist}\pi_{E2}) = \mathcal{G}(ZH^j_{dist}), \mu = H^j_{dist}, \sigma_{dist}; \tag{4.20}$$

$$P(ZH^j_{dir}|H^j_{dir}\pi_{E2}) = \mathcal{G}(ZH^j_{dir}), \mu = H^j_{dir}, \sigma_{dir}; \tag{4.21}$$

$$P(ZX^j_{dist}|X^j_{dist}\pi_{E2}) = \mathcal{G}(ZX^j_{dist}), \mu = X^j_{dist}, \sigma_{dist}; \tag{4.22}$$

$$P(ZX^j_{dir}|X^j_{dir}\pi_{E2}) = \mathcal{G}(ZX^j_{dir}), \mu = X^j_{dir}, \sigma_{dir}; \tag{4.23}$$

$$P(ZY^j_{dist}|Y^j_{dist}\pi_{E2}) = \mathcal{G}(ZY^j_{dist}), \mu = Y^j_{dist}, \sigma_{dist}. \tag{4.24}$$

$$P(ZY^j_{dir}|Y^j_{dir}\pi_{E2}) = \mathcal{G}(ZY^j_{dir}), \mu = Y^j_{dir}, \sigma_{dir}. \tag{4.25}$$

In a real-world application, the standard deviation for each sensor model choice would be related to the quality of detection. When vision is applied to detect predator, prey and home, each one can be characterised by a different colour. Consequently, the same image processing method can be employed to treat image frames, but it is possible that the observation for one colour is more precise than for another. In addition, in the environment considered, it can happen that several other objects have one of the chosen colours and the miss-detection rate is bigger for this specific colour. In the simulation, these aspects were not modelled, so standard deviations for all sensor models are constant and equal for all observations.

Next in the joint distribution, there are the motor command models. The rotation speed is rather complex, as it depends on all state variables. If the predator is close, then the rotation speed is chosen to escape from the predator. When the predator is far away, the robot follows the direction of the prey, if it is close. Otherwise, it follows the direction of home. If nothing is near enough, the rotation speed is around zero, so the robot always moves straight ahead.

The robot reacts to the predator's presence when the distance is three or less. In relation to prey and home, the robot reacts when the distance is two or less. The distribution is chosen to be a normal curve, with a mean that is a function of the distance and direction of predator, prey and home, and the standard deviation is constant. If the predator distance is bigger than three, the prey distance is bigger than two, and the home distance is two or less, then the rotation speed is chosen according to the home

direction. With this knowledge about robot behaviour, we obtain the following model for rotation speed:

$$P(M^j_{rot}|S^j_{E2}\pi_{E2}) = \mathcal{G}(M^j_{rot}), \mu = f(S^j_{E2}), \sigma_{rot},$$
$$f(S^j_{E2}) = \begin{cases} H^j_{dir}, \text{if } H^j_{dist} \leq 2, Y^j_{dist} > 2, X^j_{dist} > 3\ ; \\ Y^j_{dir}, \text{if } Y^j_{dist} \leq 2, X^j_{dist} > 3; \\ X^j_{dir} - 180, \text{if } X^j_{dist} \leq 3 \text{ and } X^j_{dir} \geq 0; \\ X^j_{dir} + 180, \text{if } X^j_{dist} \leq 3 \text{ and } X^j_{dir} < 0; \\ 0, \text{Otherwise.} \end{cases} \tag{4.26}$$

The translation speed is simpler, as it depends on distance states only. If the predator is close, the translation speed is larger (to escape faster). When the predator is distant, the translation speed is limited: it decreases when the prey is very close. If the prey is also distant, the translation speed decreases in relation to the home distance: when it is very close to home, it stops, assuming it is at home. If nothing is close enough, the translation speed is made a default value, different from zero.

$$P(M^j_{trans}|Y^j_{dist}X^j_{dist}H^j_{dist}\pi_{E2}) = \mathcal{G}(M^j_{trans}), \mu = f(Y^j_{dist}X^j_{dist}H^j_{dist}), \sigma_{trans},$$
$$f(Y^j_{dist}X^j_{dist}H^j_{dist}) = \begin{cases} 0, \text{if } H^j_{dist} = 0, Y^j_{dist} > 2, X^j_{dist} > 3\ ; \\ 1, \text{if } H^j_{dist} = 1, Y^j_{dist} > 2, X^j_{dist} > 3\ ; \\ 1, \text{if } Y^j_{dist} \leq 1, X^j_{dist} > 3; \\ 3, \text{if } Y^j_{dist} = 2, X^j_{dist} > 3; \\ 4, \text{if } X^j_{dist} \leq 3; \\ 2, \text{Otherwise.} \end{cases} \tag{4.27}$$

The last line of the joint distribution includes the probability distributions for the initial conditions of all related variables.

Prediction for this example consists of calculating probability distributions over all variable states:

$$P(S^t_{E2}|z^{0:t-1}_{E2}m^{0:t-1}_{trans}m^{0:t-1}_{rot}\pi_{E2}).$$

Adding observations, the estimation question in this second example is:

$$P(S^t_{E2}|z^{0:t}_{E2}m^{0:t-1}_{trans}m^{0:t-1}_{rot}\pi_{E2}).$$

The motor question becomes:

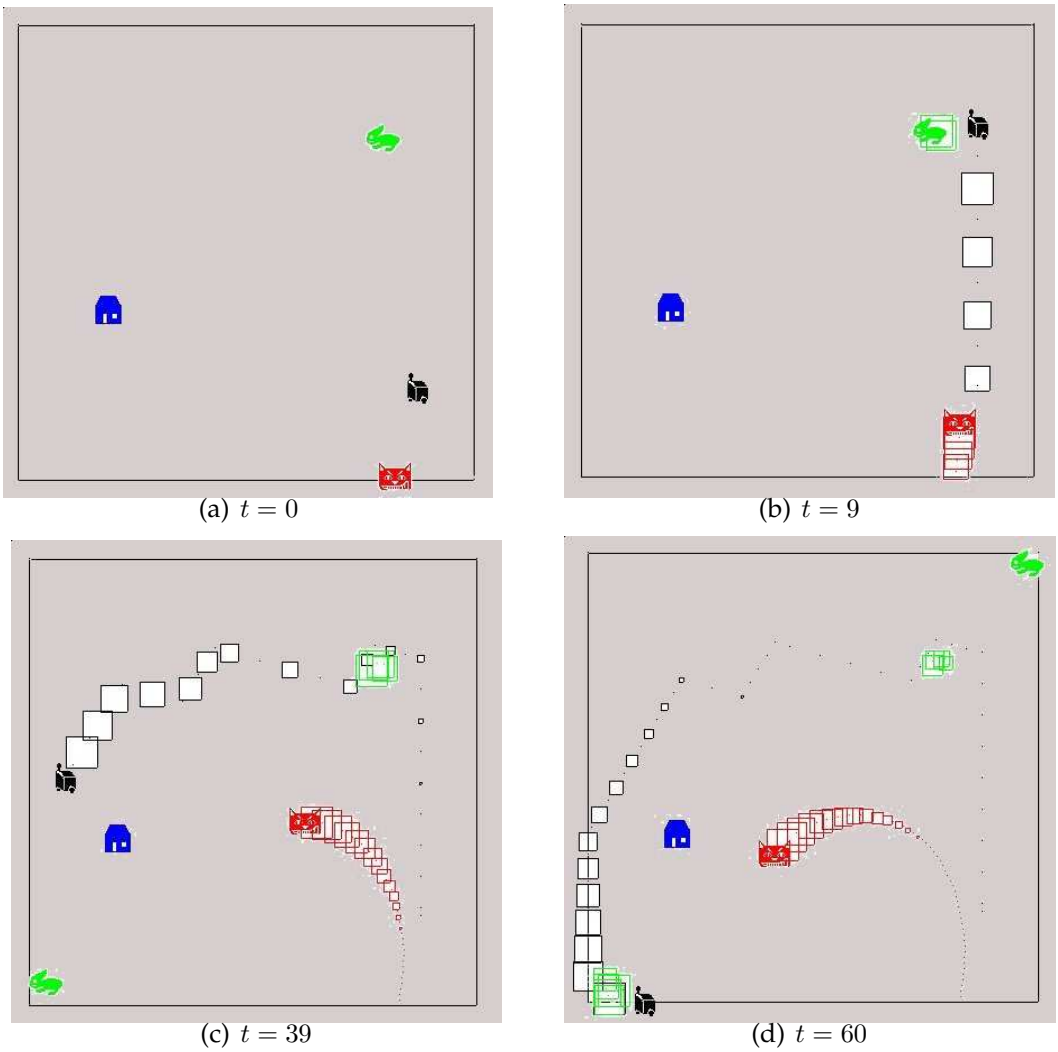
$$P(M^t_{trans}M^t_{rot}|z^{0:t}_{E2}m^{0:t-1}_{trans}m^{0:t-1}_{rot}\pi_{E2}).$$

Simulation Results

In Figure 4.6, a sequence of snapshots is shown to illustrate a simulation run. It can be seen that the robot initially escapes from the predator, and, only when the predator

is considered to be distant, chases a nearby prey. Prey capture happens as shown in Figure 4.6(c): the simulation environment destroys the caught prey and creates another one in a distant position.

After capturing a prey, the robot continues to escape from the approaching predator, and it comes near home. It begins to follow the home direction, and actually approaches a new prey: as the prey is not far, the robot ignores home, follows the prey direction, and captures it, as shown in Figure 4.6(d).



(a) $t = 0$ (b) $t = 9$

(c) $t = 39$ (d) $t = 60$

Figure 4.6: Sequence of snapshots for the simulation of Example 2.

4.1.4 Additional Main Assumptions

- Present-time state variables depend on both previous-time states and previous-time motor variables.

This assumption is mathematically expressed by the dynamic model, which describes the dynamics of the state variables.

- First-order Markov assumption

  This assumption concerns time depth considerations when writing relations between variables. A Bayes filter considering the whole dependence on past-time variables would result in the following joint distribution:

$$
\begin{aligned}
&P(S^{0:t}Z^{0:t}M^{0:t}|\pi_f) \\
&= \left| \begin{array}{l}
P(S^0Z^0M^0|\pi_f) \\
\times P(S^1|S^0M^0\pi_f) \times P(Z^1|S^{0:1}\pi_f) \times P(M^1|S^{0:1}M^0\pi_f) \\
\times P(S^2|S^{0:1}M^{0:1}\pi_f) \times P(Z^2|S^{0:2}\pi_f) \times P(M^2|S^{0:2}M^{0:1}\pi_f) \\
\dots \\
\times P(S^t|S^{0:t-1}M^{0:t-1}\pi_f) \times P(Z^t|S^{0:t}\pi_f) \times P(M^t|S^{0:t}M^{0:t-1}\pi_f).
\end{array} \right.
\end{aligned}
$$

  It is possible, however, to define a limit in time depth at t_1, as in:

$$P(S^t|S^{t_1:t-1}M^{t_1:t-1}\pi_f).$$

  Actually, the Markov assumption imposes a limit to time depth, and a first-order Markov assumption assigns the time depth to be equal to 1.

  Assuming our filter is a first-order Markov system, the dynamic model is written as $P(S^t|S^{t-1}M^{t-1}\pi_f)$ and the motor model becomes $P(M^j|S^jM^{j-1}\pi_f)$.

- The dynamic, sensor and motor models are stationary in time.

  Considering time depth as defined by the first-order Markov assumption, if no stationarity is assumed, we have the filter joint distribution:

$$
\begin{aligned}
&P(S^{0:t}Z^{0:t}M^{0:t}|\pi_f) \\
&= \left| \begin{array}{l}
P(S^0Z^0M^0|\pi_f) \\
\times P(S^1|S^0M^0\pi_f) \times P(Z^1|S^1\pi_f) \times P(M^1|S^1M^0\pi_f) \\
\times P(S^2|S^1M^1\pi_f) \times P(Z^2|S^2\pi_f) \times P(M^2|S^2M^1\pi_f) \\
\dots \\
\times P(S^t|S^{t-1}M^{t-1}\pi_f) \times P(Z^t|S^t\pi_f) \times P(M^t|S^tM^{t-1}\pi_f).
\end{array} \right.
\end{aligned}
$$

  That would mean that, in the observation model, for example, the relation between observations and states changes with time. The stationarity assumption, often employed in robotics, simplifies the problem considerably:

$$P(Z^0|S^0\pi_f) = P(Z^1|S^1\pi_f) = P(Z^2|S^2\pi_f) = P(Z^t|S^t\pi_f). \tag{4.28}$$

	$N_s = N_z = 1$	$N_s = N_z = 2$	$N_s = N_z = 4$	$N_s = N_z = 16$
$N_m = 1$	10^7	10^{11}	10^{15}	10^{24}
$N_m = 2$	10^9	10^{12}	10^{16}	10^{24}
$N_m = 4$	10^{12}	10^{14}	10^{17}	10^{24}
$N_m = 16$	10^{20}	10^{20}	10^{22}	10^{26}

Table 4.1: Approximate number of probability values to be stored as a function of the number of states, motor and observation variables. All variables are considered to have the same number of possible values, noted n.

4.2 Requirements Analysis

4.2.1 Space Complexity

For comparison with the first proposed method to answer the ultimate question, we present in Table 4.1 the memory space necessary for storing the Bayes filter proposed in this chapter. In this table, the number of state variables is considered to be equal to the number of observed variables.

Comparing with Table 1.1, a substantial reduction in memory requirement is noticeable: in particular, it does not matter how long the Bayes filter is used, the amount of memory stays constant. It can be seen, however, that the amount of memory still increases greatly with the number of variables.

It is then interesting to verify how the amount of necessary memory space grows with the number of variables. For the Bayes filter proposed in this chapter, we must keep in memory one table for each term in the joint distribution necessary to perform the calculation of the prediction question (equation 4.9), the estimation question (equation 4.6) and the motor question (equation 4.8). Below, the expression of the size for one table for each term is developed.

- The term $P(S^j|S^{j-1}M^{j-1}\pi_f)$ is used in the prediction question and it can be stored in a table of size $n^{2N_s+N_m}$.
- The previous time step estimation, $P(S^{t-1}|z^{0:t-1}m^{0:t-2}\pi_f)$, is also necessary in the prediction question, and it is a table of size n^{N_s}.
- The prediction question results in a table of size n^{N_s}, to be used in calculating the estimation question.
- The sensor model term $P(Z^j|S^j\pi_f)$, used in the estimation question, is stored in a table of size $n^{N_s+N_z}$.
- Estimation question results are stored in a table of size n^{N_s}.
- The term $P(M^j|S^jM^{j-1}\pi_f)$ employed in the motor question has a size of $n^{N_s+2N_m}$.

- The motor question result is stored in a table of size n^{N_m}.

The number of cells required to store all necessary terms in memory is:

$$\mathcal{M} = n^{2N_s+N_m} + 3n^{N_s} + n^{N_s+N_z} + n^{N_s+2N_m} + n^{N_m}. \tag{4.29}$$

Applying the calculation in Equation 4.29 to Example 2 presented in this chapter, we find that nearly 1.2×10^{10} probability values must be stored. This number is approximate: the number of variables and the number of cases of each variable in the example were applied in the general form of the joint distribution (as shown in equation 4.1). The number of probability values to be stored can be reduced if particular choices of conditional independence considered in the example joint distribution (equation 4.15) are taken into account. We adopt the general joint distribution to use this result as an evaluation parameter between different version of filters applied to the same example.

The approximative number of 1.2×10^{10} probability values is mostly because of the exponential dependence on the number of state variables, in both the dynamic and sensor models. It is important to notice that this number does not increase with time: it is fixed for utilisation independent of the number of time steps.

The expression in equation 4.29 has an upper bound of:

$$\mathcal{M} = O(n^{2N_s+N_m} + n^{N_s+N_z} + n^{N_s+2N_m}). \tag{4.30}$$

The green part of the Equation 4.30 (the first term) is related to the amount of memory needed to store the dynamic model. Its double exponential growth rate is caused by the time dependence between state variables.

This dependence is also present in the motor model, and consequently the brown term in the equation above, related to the motor model, shows a double exponential rate with the number of motor command variables.

The blue term in the above equation relates to the sensor model: it is considered here to be one big sensor model relating all observation variables to all state variables. We saw in the previous chapter (Section 3.2.1) that sensor fusion schema can reduce significantly the amount of space required for sensor models: instead of $n^{N_s+N_z}$, only $N_z n^{N_s+1}$ probability values would be necessary. The sensor fusion schema modifies the dependence on the number of observation variables from exponential to linear:

$$\mathcal{M} = O(n^{2N_s+N_m} + N_z n^{N_s} + n^{N_s+2N_m}). \tag{4.31}$$

The main problem in Bayes filter utilisation is the exponential dependence on the number of state variables. In robotics applications, the number of state variables is directly related to the difficulty of the specified task and the environment sophistication: the number of state variables increases with task complexity and with rich environments.

The number of observation variables is often bigger than the number of state variables and, in this case, utilisation of sensor fusion helps to reduce the dependence on the number of observation variables from exponential to linear.

4.2.2 Time Complexity

For robotics applications, time complexity is a delicate issue, as it influences the ability to fulfil real-time constraints.

The computational effort for the Bayes filter is the sum of operations necessary for solving the three questions at each time step: first, the prediction equation as shown in equation 4.9; then, estimation according to equation 4.6; and finally, the motor question as stated in equation 4.8.

Here, addition and multiplication are considered atomic operations with equal costs. Normalisation constant calculations are also neglected, as it is always possible to normalise the calculated probability distribution later.

- The prediction question, in equation 4.9 and rewritten below, involves nested sums over the state space:

$$\begin{aligned} & P(S^t|m^{0:t-1}z^{0:t-1}\pi_f) \\ & \propto \sum_{S^{t-1}} \left[\begin{array}{l} P(S^t|S^{t-1}m^{t-1}\pi_f) \\ \times P(m^{t-1}|S^{t-1}m^{t-2}\pi_f) \\ \times P(S^{t-1}|z^{0:t-1}m^{0:t-2}\pi_f) \end{array} \right] . \end{aligned}$$

  For each value of the state set S^{t-1}, it is necessary to perform two multiplications, which gives a total of $2n^{N_s}$ multiplications. Then, these values must be summed: $n^{N_s} - 1$ sums are required. These calculations are executed for each value of state set S^t, and the total number of operations for answering the prediction question is thus $3n^{2N_s} - n^{N_s}$.

- On the other hand, the estimation question, as stated in equation 4.6 and repeated below, requires only one multiplication of terms for each value of the state set S^t, resulting in n^{N_s} operations.

$$\begin{aligned} & P(S^t|z^{0:t}m^{0:t-1}\pi_f) \\ & \propto \left| \begin{array}{l} P(z^t|S^t\pi_f) \\ \times P(S^t|m^{0:t-1}z^{0:t-1}\pi_f) \end{array} \right. . \end{aligned}$$

- The motor control question, in equation 4.8 and repeated below, shows a summation over the state space.

$$\begin{aligned} & P(M^t|z^{0:t}m^{0:t-1}\pi_f) \\ & \propto \sum_{S^t} \left[\begin{array}{l} P(M^t|S^t m^{t-1}\pi_f) \\ \times P(S^t|m^{0:t-1}z^{0:t}\pi_f) \end{array} \right] . \end{aligned}$$

It requires one multiplication for each value of the state set for a total of n^{N_s} multiplications plus $n^{N_s} - 1$ additions. These calculations are necessary for each value of the motor set M^t, which implies $2n^{N_m+N_s} - n^{N_m}$ operations.

Adding all the above operations, the total number is equal to:

$$\mathcal{C} = 3n^{2N_s} + 2n^{N_m+N_s} - n^{N_m}. \tag{4.32}$$

In our illustrative example, approximately 1.4×10^8 operations are required to execute all the calculations at each time step. This number was calculated employing general forms of question expressions, and not considering conditional independence.

Particular choices of conditional dependence can decrease this number, but it is not taken into account here so that comparisons with other filters applied to the same example are possible.

The number of mathematical operations necessary for solving the three questions then has an upper bound of:

$$\mathcal{C} = O(n^{2N_s} + n^{N_m+N_s}) \tag{4.33}$$

Similar to space complexity, we find again an exponential dependence on the number of state variables. The green term in the above equation corresponds to the prediction question, and the double rate of exponential grow is because of the summation over the state variables from the previous time step. In other words, it is caused by the propagation of uncertainty about the state variables, from time $t-1$ to time t.

The brown term is the consequence of asking the motor question, and can be seen as the cost of selecting a subset of variables for taking a decision over their values.

It is interesting to note that the observation variables have no influence on the number of operations necessary, as only one sensor model term is considered in the equations. If the sensor fusion method is employed, N_z models, one for each observation variable, would appear in the estimation question expression, and the number of mathematical operations would be:

$$\mathcal{C} = 2n^{2N_s} + (N_z - 1)n^{N_s} + 2n^{N_m+N_s} - n^{N_m}. \tag{4.34}$$

The upper bound is:

$$\mathcal{C} = O(n^{2N_s} + N_z n^{N_s} + n^{N_m+N_s}). \tag{4.35}$$

Sensor fusion results in a linear dependence on the number of observation variables (the blue term in the above equation). In practical applications, where the number of observation variables is often bigger than the number of state variables, this linear dependence is still less important that the double exponential dependence on the number of state variables. As explained in the previous chapter, the decision to use sensor fusion is a trade-off between increased calculation time and memory space reduction, together with simplifications in programming.

4.2.3 Programming

It is necessary to program three terms: the dynamic model, the sensor model and the motor control.

These terms have precise meanings, directly related to the robot, environment and task being programmed. For this reason, even if a table is employed instead of parametric forms, each value can be associated with a real situation, and can be selected accordingly.

Learning

The Bayes filter presented in this chapter can benefit from the same strategies and methods proposed in the last chapter for learning the sensor model.

In addition, learning approaches applied to learning Markov processes can be adapted and extended to learn both dynamic and sensor models. Murphy [2002b] gives a good introduction to such methods. Particularly popular are the steepest ascent in likelihood technique for learning Bayesian networks [Binder et al., 1997]; and the expectation maximisation method used for learning both the sensor and dynamic models on a hidden Markov model [Rabiner, 1989]. Originally proposed for speech recognition applications, this method was extended to robotics applications [Koenig and Simmons, 1996; Shatkay and Kaelbling, 1997; Thrun et al., 1998; Dellaert et al., 2003].

The state aggregation method was proposed by Nikovski [2002] and it explores an alternative approach of merging percept trajectories.

4.3 Discussion

4.3.1 Robotics

Probabilistic Approaches in Robotics

Several probabilistic approaches have been applied in robotics, and they differ basically by the number and type of constraints and assumptions. Bayes rule-based approaches can be grouped and arranged in a tree-like graph, as initially proposed by Murphy [2002a] and extended by Diard [2003]. The extended version of this tree- like graph is shown in Figure 4.7.

In this graph, an approach is considered more general when it is based on a smaller number of assumptions or constraints. As assumptions and constraints are added, approaches become more specific and special cases of more general ones.

The Bayes filter as presented by Diard [2003] (also called Bayesian filter or Bayesian tracking [Arulampalam et al., 2002]), is related to dynamic Bayesian networks with only state and observation variables. It is composed of an a priori term over states, a dynamic (or transition) model and a sensor (or observation) model. As can be seen in Figure 4.7, it is an specialisation of dynamic Bayesian networks.

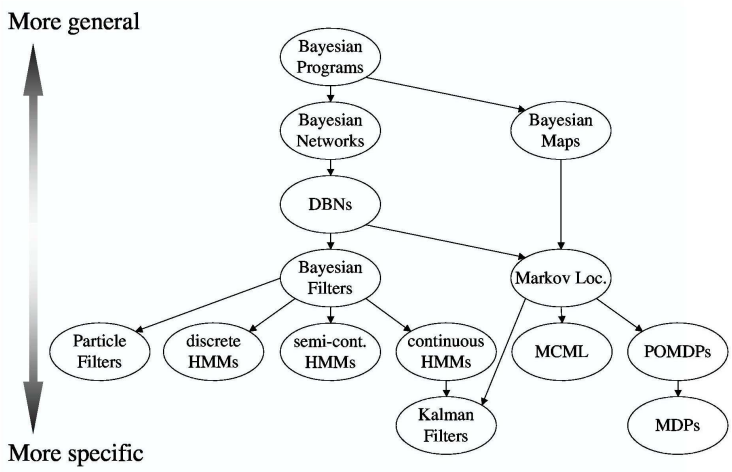


Figure 4.7: Relationships between Bayesian approaches, from Diard [2003].

Markov localisation, also called an input-output hidden Markov model, is also a specialisation of dynamic Bayesian networks. It has the additional influence of an action variable in the dynamic model and it has been widely applied in robotics [Fox et al., 1998b; Thrun, 2000].

The Bayesian Program described in this chapter is not classified as a specialisation of Markov localisation because of the presence of the motor model in the joint distribution. We place our proposal as a kind of Bayes filter. Motor variables, not included in the general structure of a Bayes filter, are employed in our proposal, but they are very similar to state variables. Motor variables have associated dynamics, like state variables, given by the motor model. Their values are also estimated taking into account observations from the environment: the state estimation expression employed in the motor question solution includes observations.

Actually, motor variables are different from state variables because they are not used in the sensor model, and because they have known values. At the end of each iteration, a decision is taken on the motor command values and at the next iteration, the motor variables have attributed values, while state variables are never precisely known, but only estimated.

Time Horizon for Decision

Partially Observable Markov Decision Processes, known as POMDP, are an extension of Markov localisation, as one additional constraint is added in relation to action decisions [Diard et al., 2003]. In contrast to Markov localisation, where no assumption is included about how actions are decided, POMDP proposes to assign a reward value to each possible action to compare the adequacy of actions. This reward value is associated with executing a given action in a given state situation. A policy is produced which prescribes the choice of action for any possible state [Cassandra et al., 1994; Simmons and

Koenig, 1995; Kaelbling et al., 1998]. Finding the policy involves maximising reward in a finite time horizon and it is very hard to be calculated.

In our proposal, the motor model is responsible for assigning the relationship between states and motor commands, and in the answer to the motor question, it is weighted by the state estimation. The result of the motor question associates a probability value with each possible motor command according to the current state estimation.

While POMDP explicitly considers future time when evaluating a sequence of actions for maximising reward, in our filter decisions take only the present time into account. Another difference is that the reward function in POMDP can consider other variables as well as state variables, while in our motor model only state variables influence motor command decisions.

Our proposal for action decision can be seen as a simpler alternative to POMDP, especially useful when fast decisions are essential.

In highly dynamic environments, it is necessary to take decisions on actions as often as possible. Planning (searching the future for the best action sequence) has two major drawbacks: first, it takes a long time to build an optimised plan; second, the world is changing continuously, and any plan may become outdated very quickly. Our approach aims to take decisions frequently: these decisions may not be optimal, but in due time, they can cope with a fast-changing world.

Motor Variables are State Variables

As discussed above, our filter is a kind of Bayes filter, as the motor variables are like state variables with known values: present-time motor variables are similar to state variables, and the motor question estimates its values, as shown in equation 4.8.

It is possible to see in the time complexity expression in equation 4.33 that one exponential term results from asking the motor question, which is necessary for deciding motor command values.

When motor variables are considered fully as state variables, the motor question is not necessary and this term in time complexity disappears. In this case, however, the motor variables would be added to the state variables in the prediction and estimation question, and consequently the uncertainty regarding their values would be propagated recursively from time step to time step. In this situation, time complexity would be:

$$\mathcal{C} = O(n^{2(N_s+N_m)}) \tag{4.36}$$

Thus, the influence of motor command decisions on complexity is twofold: first, asking the motor question adds an exponential term (the brown one in equation 4.33); second, once values are decided, uncertainty regarding motor command values is not propagated to the next time step, and the double exponential term only relates to the number of state variables (the green term in equation 4.33).

If it is desired to decide the values of motor commands, the motor question is necessary. Once their values are decided, they can be ignored, and the associated uncertainty

can be propagated. In this situation, the complexity is similar to equation 4.36.

4.3.2 Cognition

The cognitive interpretations of the mathematical hypotheses added in this chapter (see decomposition in equation 4.1) are the following:

- The Central Nervous System uses a dynamic model to predict the evolution of its internal states based on the preceding internal states and on the preceding motor commands.
$$P(S^j|S^{j-1}M^{j-1}\pi_f)$$
- It also uses a motor model to decide of motor command according to internal states and previous motor commands.
$$P(M^j|S^jM^{j-1}\pi_f)$$
- Short past time horizon is sufficient for both these models (first-order Markov assumption).
- These models are stationary: they do not evolve with time.

The different steps to answer the ultimate question are:

- Predict your internal state as a weighted sum of the different dynamical models, where the weights are the estimated probability on the past internal states updated by the just chosen motor command (equation 4.9).
- Update this prediction by taking into account the sensory inputs to estimate the probability of the present internal states (equation 4.6).
- Decide the new motor commands as a weighted sum of the different motor models, where the weights are the just updated estimated probability of the present internal states (equation 4.8).

This Bayes filter mechanism is a compromise between: (i) the need to take into account the history of sensory inputs and motor decisions to solve the ultimate question, and, (ii) the impossibility to memorise and process the great quantity of information involved.

It is proposed to summarise all the past information in a probability distribution on possible internal states, alternatively predicted and estimated, independent on how long is the history of these past informations. A major consequence is that space and time complexity do not present the value of the time instant (the duration of history) multiplying all exponents, as it was the case for solution shown in Section 1.1. This

simplification is even more drastic than the one obtained in the previous chapter after introducing internal states.

A cognitive consequence of the Bayes filter mechanism presented is that the Central Nervous System would work with probability distributions on internal states. Taking a decision about internal states means using the decided values instead of the probability distributions: this is equivalent to losing all past information and destroying the Bayes filter recursive estimation utility.

Beyond the success of Bayesian models in computational neuroscience this important consideration justifies the interest in this area of research. To consider that the Central Nervous System processes probability distributions instead of values is a very important shift in the point of view on brain studies, and it can bring important consequences on research trends.

All the fundamental questions about information coding, inference processing and memory storage have to be revisited with this different conception of the whole system. Now, the parameters defining the probability distributions have to be coded, processed and stored in a coherent, coordinated and synchronous manner, most likely distributed in time, in space or in both.

In Bayesian reasoning, taking a decision means to replace a complete probability distribution by a single value, therefore losing information. For this reason, taking a decision over one variable value is undesired from an informational point of view. But from a processing point of view, decisions lead to simplifications, essentially by suppressing some of the required integration. To decide or not to decide at some point about one variable may be seen as a trade off between keeping all the information about this variable and having better tractable computation.

Sensory-motor systems impose a decision on motor commands at each time step. Each of these decisions is a commitment. This motor commitment "simplifies" the world: perception is oriented, fewer acting options are opened for the future, and reasoning is more tractable. However, if the commitment is unsuitable because inadequate motor commands have been decided, then a lot of efforts are necessary to take the correcting actions. In some situations, it may happen that it is very difficult to realise that wrong decisions have been taken, as perception itself is strongly oriented by the wrong decision (second line of equation 4.5).

Taking a decision is considered as a rational evaluation of the risk and utility of possible actions, but at lower cognitive levels, decision is also intimately related to perception. In the presence of certain visual stimuli, an animal must choose between escape and chase: there is no time for evaluating costs and benefits, and a decision must be taken rapidly. Deliberation and plan-making are unavoidable for some situations, but in other circumstances they are undesirable. We support the idea that the time horizon when taking a decision must be adapted to each particular situation.

Stationary models seen from a cognitive point of view may at first be very disturbing: any learning that could change models in time seems to be excluded. It is important to discuss this hypothesis considering that different time constants are involved. Motor decision and internal state updating requires time constants on the order of a tenth of a

second. Time constant for learning motor or perception abilities is rather on the order of minutes, hours or days. Consequently, learning may take place, yet preserving the hypothesis of stationary models at a time scale of a tenth of a second.

4.4 Summary and Open Questions

In this chapter, we presented a Bayes filter for solving the ultimate question, breaking the dependence on time. The first-order Markov assumption ensures a limited depth when considering the past, and the stationarity assumption allows the validity of the models for all time instants. The motor model is also included directly in the filter joint distribution:

$$\begin{aligned} & P(M^{0:t} S^{0:t} Z^{0:t} | \pi_f) \\ & = \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S^j | S^{j-1} M^{j-1} \pi_f) \\ \times P(Z^j | S^j \pi_f) \\ \times P(M^j | S^j M^{j-1} \pi_f) \end{array} \right] \\ \times P(M^0 S^0 Z^0 | \pi_f). \end{array} \right. \end{aligned}$$

Utilisation consists of asking a sequence of questions, and the recursive use of previous time step answers reduce the amount of calculation necessary.

The prediction expression uses the estimation from the previous time step:

$$\begin{aligned} & P(S^t | m^{0:t-1} z^{0:t-1} \pi_f) \\ & \propto \sum_{S^{t-1}} \left[\begin{array}{l} P(S^t | S^{t-1} m^{t-1} \pi_f) \\ \times P(m^{t-1} | S^{t-1} m^{t-2} \pi_f) \\ \times P(S^{t-1} | z^{0:t-1} m^{0:t-2} \pi_f) \end{array} \right] . \end{aligned}$$

The answer to the estimation question utilises the prediction expression:

$$\begin{aligned} & P(S^t | z^{0:t} m^{0:t-1} \pi_f) \\ & \propto \left| \begin{array}{l} P(z^t | S^t \pi_f) \\ \times P(S^t | m^{0:t-1} z^{0:t-1} \pi_f). \end{array} \right. \end{aligned}$$

In contrast to other Bayes filters, the motor question is now part of the recursive calculation loop:

$$\begin{aligned} & P(M^t | z^{0:t} m^{0:t-1} \pi_f) \\ & \propto \sum_{S^t} \left[\begin{array}{l} P(M^t | S^t m^{t-1} \pi_f) \\ \times P(S^t | m^{0:t-1} z^{0:t} \pi_f) \end{array} \right]. \end{aligned}$$

The proposed solution to the ultimate question considerably reduces the space and time required in comparison with the naive solution in Chapter 1.

The complexity is, however, still too great for practical implementations. The main reason is the cost of summing over the state space in the expressions answering the prediction and motor questions. The next chapter defines constraints over state and observation variables to reduce complexity.

Chapter 5

Exploiting Conditional Independence

Key Points:

- **Exploit the independence between the state and observation variables to reduce time complexity in the prediction question.**
- **Fusion with coherence is a helpful technical resource.**

Bayes filter solution proposed in previous chapter removed time dependence, applying the first-order Markov and stationarity assumptions. However, it still presents huge space and time complexity.

In this chapter we aim to reduce the complexity by exploiting conditional independence between the state and observation variables. This independence is mainly because of the existence of disjoint subsets in both the state and observation spaces.

Motor variables, however, are related to all other variables, and no independence assumption can be applied. A helpful resource for dealing with this situation is coherence-based fusion [1].

5.1 Bayesian Program

In the previous chapter, the huge calculation time was mainly because of the prediction and motor questions: both required nested sums over the state space.

In this chapter, conditional independence in the state and observation spaces is exploited to reduce the nested summations. The state space is preserved, and summations are still executed, but they can now be calculated independently.

While conditional independence can exist in the state and observation spaces, it cannot be applied to the motor command variables. This creates a problem when defining the joint distribution, as shown below:

[1] This technique was proposed by Pradalier et al. [2003b,a] where it was named fusion with diagnosis. In our work, the *diagnosis variable* is semantically nearer to a coherence variable, and we take the liberty of calling it *coherence-based fusion*.

$$
\begin{aligned}
&P(M^{0:t}S^{0:t}Z^{0:t}|\pi_f) \\
&= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_{lf}) \\ \prod_{i=1}^{N_i} P(Z_i^j|S_i^j\pi_{lf}) \\ \prod_{i=1}^{N_i} P(M^j|S_i^jM^{j-1}\pi_{lf}) \end{array} \right] \\ \times P(M^0S^0Z^0|\pi_{lf}). \end{array} \right.
\end{aligned}
$$

The definition of the motor model as

$$\prod_{i=1}^{N_i} P(M^j|S_i^jM^{j-1}\pi_{lf})$$

is not correct according to the rules detailed in chapter 2: variable M^j appears more than once on the left. However, a product of motor models can be achieved using intermediate variables λ_i^j, called coherence variables. The motor model definition becomes:

$$P(M^j|\pi_{lf}) \prod_{i=1}^{N_i} P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_{lf}).$$

Expressing a model using coherence variables is called coherence-based fusion, and it does not add new knowledge nor new assumptions. This method is employed in this chapter for defining the motor model in the proposed Bayesian program.

5.1.1 Description

Additional Relevant Variables

A new type of variable is employed to indicate whether the set of concerned variables in the model present coherent values. In this chapter, λ^t is a set of binary coherence variables for the motor command model.

Decomposition

$$
\begin{aligned}
&P(M^{0:t}S^{0:t}Z^{0:t}\lambda^{0:t}|\pi_{lf}) \\
&= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^j|S_i^{j-1}M^{j-1}\pi_i)\right] \\ \prod_{i=1}^{N_i} \left[P(Z_i^j|S_i^j\pi_i)\right] \\ \times P(M^j|\pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i)\right] \end{array} \right] \\ \times P(M^0S^0Z^0\lambda^0|\pi_{lf}). \end{array} \right.
\end{aligned}
\tag{5.1}
$$

The first line inside the time product is the dynamic model, which is decomposed as the product of N_i terms, each one related to a subset S_i of the state variables. To obtain this product, N_i subsets must be mutually exclusive, so that:

$$S^j = \bigcup_{i=1}^{N_i} [S_i^j], \bigcap_{i=1}^{N_i} [S_i^j] = \emptyset, \forall j. \tag{5.2}$$

Each state subset $S_i, i = 1, \ldots, N_i$ is composed of N_{s_i} state variables, so that $\sum_{i=1}^{N_i} N_{s_i} = N_s$ and for each subset, it is possible to define a dynamic model.

The second line in the joint distribution is the sensor model. It is also decomposed as a product of N_i terms, and each term uses the same state subsets defined above. Consequently, the observation variable space is also split into N_i mutually exclusive subsets:

$$Z^j = \bigcup_{i=1}^{N_i} [Z_i^j], \bigcap_{i=1}^{N_i} [Z_i^j] = \emptyset, \forall j. \tag{5.3}$$

In the third line, the motor command model is now defined using coherence-based fusion. The second term in this line defines the model of coherence between state subsets and motor commands, and the coherence variables λ are defined as being assembled into N_i mutually exclusive subsets:

$$\lambda^j = \bigcup_{i=1}^{N_i} [\lambda_i^j], \bigcap_{i=1}^{N_i} [\lambda_i^j] = \emptyset, \forall j. \tag{5.4}$$

In the last line of the joint distribution, outside the time products, are the initial conditions distributions.

The joint distribution in equation 5.1 defines a filter called a *global filter*, and it is defined over all relevant variables. Following the partition of the variable space into the mutually exclusive subsets shown in equations 5.2, 5.3 and 5.4, it is possible to derive the existence of *elementary filters*, one for each variable subset. The demonstration of the equivalence between utilisation of global or elementary filters is available in Appendix B.

The joint distribution of one general elementary filter is then defined as:

$$\begin{aligned} & P(M^{0:t} S_i^{0:t} Z_i^{0:t} \lambda_i^{0:t} | \pi_i) \\ & = \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i). \end{array} \right. \end{aligned} \tag{5.5}$$

Inside the time product, we can see that the terms are now related to only one variables subset. The first line contains the dynamic model on the state variables subset S_i. In the second line, the sensor model is related to the observation variable subset Z_i and the state variables subset S_i.

The third line shows the motor model relating the state variables subset S_i and the motor variables M. It is defined here using coherence-based fusion. The relation

between the inverse motor model [2] and the coherence- based fusion motor model is very simple. Considering the inverse motor model given as $P(M^j|S_i^j M^{j-1}\pi_i)$, we have for all possible values of M^j, S_i^j and M^{j-1}:

$$\begin{aligned} P([\lambda_i^j = 1]|M^j S_i^j M^{j-1}\pi_i) &= P(M^j|S_i^j M^{j-1}\pi_i) \\ P([\lambda_i^j = 0]|M^j S_i^j M^{j-1}\pi_i) &= 1 - P(M^j|S_i^j M^{j-1}\pi_i). \end{aligned} \tag{5.6}$$

In the last line of the elementary filter joint distribution, we can see the probability distribution for the initial conditions of variables relevant to this elementary filter.

Bayesian network representations of both elementary and global filter joint distributions are presented in appendix B, Figures B.1 and B.2, respectively.

5.1.2 Utilisation

Instead of asking questions to a global joint distribution, we now wish to use elementary filters as much as possible. From the diagram shown in Figure 5.1, it can be seen that the basic utilisation of an elementary filter i is similar to the utilisation shown in the previous chapter: prediction, estimation and motor questions can be easily identified.

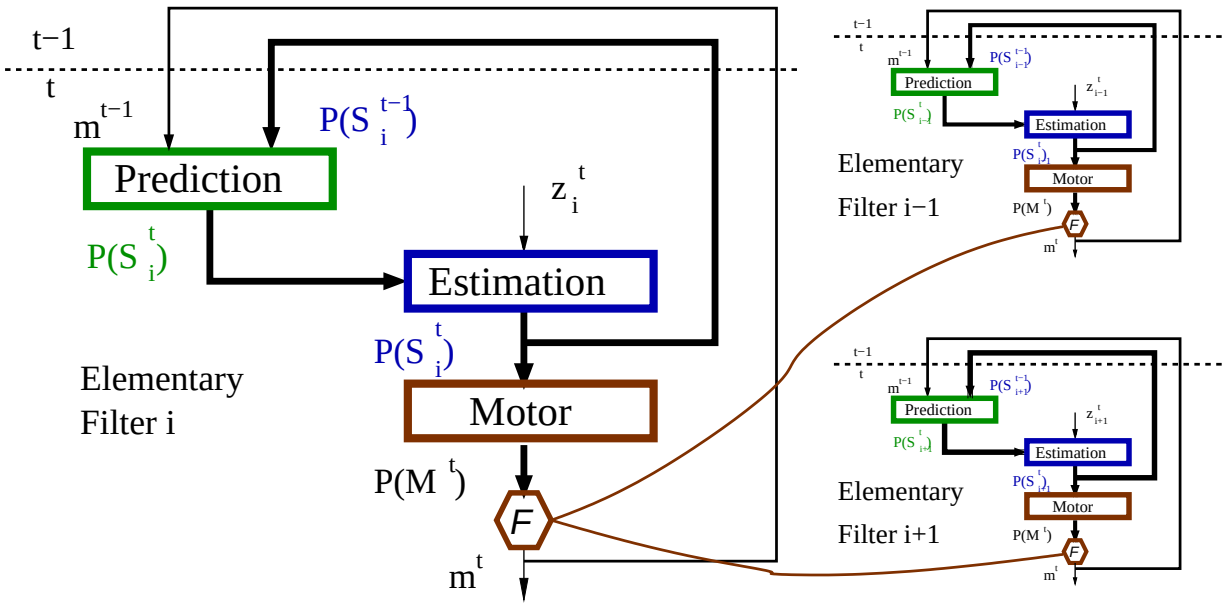


Figure 5.1: Utilisation of elementary filters.

It is demonstrated in Appendix B, that asking prediction, estimation and motor question to the global filter is analogous to asking prediction, estimation and motor questions to each elementary filter. Here, we show the results of this demonstration and briefly discuss its characteristics.

[2] *Inverse model* in the sense described by Pradalier et al. [2003b]

Prediction in an Elementary Filter

$$\begin{array}{l} P(S_i^t|m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}\pi_i) \\ \quad \propto \sum_{S_i^{t-1}} \left[\begin{array}{l} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\ \times P(m^{t-1}|\pi_i) \times P(\lambda_i^{t-1}|m^{t-1}S_i^{t-1}m^{t-2}\pi_i) \\ \times P(S_i^{t-1}|m^{0:t-2}z_i^{0:t-1}\lambda_i^{0:t-2}\pi_i) \end{array} \right]. \end{array} \tag{5.7}$$

The expression solving the prediction question in each elementary filter is very similar to the prediction expression for the Bayes filter shown in the previous chapter (equation 4.9). The differences concern the motor model, here defined with the help of coherence-based fusion, and summing over the state subset instead of the whole state space.

Prediction in the Global Filter

$$\begin{array}{l} P(S^t|m^{0:t-1}z^{0:t-1}\lambda_i^{0:t-1}\pi_{lf}) \\ \quad \propto \prod_{i=1}^{N_i} P(S_i^t|m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}\pi_i) \quad . \end{array}$$

As variables subsets are mutually exclusive, global prediction is proportional to the product of elementary predictions. This important relation is responsible for the reduction of time complexity.

Estimation in an Elementary Filter

$$\begin{array}{l} P(S_i^t|m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}\pi_i) \\ \quad \propto \left| \begin{array}{l} P(z_i^t|S_i^t\pi_i) \\ \times P(S_i^t|m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}\pi_i). \end{array} \right. \end{array} \tag{5.8}$$

The estimation expression for an elementary filter is similar to the case of the Bayes filter presented in the previous chapter (equation 4.6), but it concerns only subsets of state and observation variables.

Estimation in the Global Filter

$$\begin{array}{l} P(S^t|m^{0:t-1}z^{0:t}\lambda^{0:t-1}\pi_{lf}) \\ \quad \propto \prod_{i=1}^{N_i} \left[\; P(S_i^t|m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}\pi_i) \; \right]. \end{array} \tag{5.9}$$

As variable subsets are mutually exclusive, global estimation is proportional to the product of elementary estimations.

Motor Question in Elementary Filter

$$\begin{aligned} & P(M^t | m^{0:t-1} z_i^{0:t} \lambda_i^{0:t} \pi_i) \\ & \propto \sum_{S_i^t} \left[\begin{array}{l} P(M^t|\pi_i) \times P(\lambda_i^t | M^t S_i^t m^{t-1} \pi_i) \\ \times P(S_i^t | m^{0:t-1} z_i^{0:t} \lambda_i^{0:t-1} \pi_i) \end{array} \right]. \end{aligned} \tag{5.10}$$

One of the main differences between the expression for the motor question in an elementary filter and the motor question expression for the Bayes filter presented in the previous chapter (equation 4.8) is the utilisation of coherence-based fusion. The coherence variable λ_i^t is made equal to one as only coherent values of motor commands are desired.

The summation over a subset of the state space, instead of over the whole state space, reduces the computational effort necessary to calculate this expression.

Motor Question in the Global Filter

$$\begin{aligned} & P(M^t | m^{0:t-1} z^{0:t} \lambda^{0:t} \pi_{lf}) \\ & \propto \prod_{i=1}^{N_i} P(M^t | m^{0:t-1} z_i^{0:t} \lambda_i^{0:t} \pi_i). \end{aligned} \tag{5.11}$$

Global motor commands result from the fusion of motor command proposals from each filter.

In the above prediction question expression, we can see that the recursive calculation loop is restricted to each elementary filter. Consequently, it is not necessary to calculate the global prediction and estimation explicitly. For the motor question, however, it is essential to have a global distribution before drawing the motor command values.

In Figure 5.1, a hexagon with letter F is included to illustrate the need to combine motor command answers. The big diagram is related to elementary filter i, but fusion is executed using information from all other filters: in the figure, filters $i-1$ and $i+1$ are shown in a smaller size, at right.

In the following chapters, whenever a combination of answers from elementary filters is necessary, it is indicated by the symbol of the hexagon with letter F. Even if no other elementary filter is shown in the figure, they are implied by the presence of fusion symbol.

5.1.3 Illustrative Example

To illustrate the definition of Elementary Filters, we return to Example 2 described in the previous chapter, with the same behaviour specification.

Bayesian Program

The joint distribution in equation 4.15 shows that an inherent independence already exists between state and observation variables. Three subsets are easily discerned: home-related variables, prey-related variables and predator-related variables. These subsets define three elementary filters with very similar structures.

Coherence variables λ and τ are added, one for each motor command variable in each elementary filter: λ for rotation speed and τ for translation speed.

The joint distribution of the home elementary filter, for example, is as follows.

$$\begin{aligned} & P(H_{dir}^{0:t} H_{dist}^{0:t} Z H_{dir}^{0:t} Z H_{dist}^{0:t} M_{rot}^{0:t} M_{trans}^{0:t} \lambda_{home}^{0:t} \tau_{home}^{0:t} | \pi_{home}) \\ = & \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(H_{dist}^{j} | H_{dist}^{j-1} \pi_{home}) \times P(H_{dir}^{j} | H_{dir}^{j-1} M_{rot}^{j-1} \pi_{home}) \\ \times P(Z H_{dir}^{j} | H_{dir}^{j} \pi_{home}) \times P(Z H_{dist}^{j} | H_{dist}^{j} \pi_{home}) \\ \times P(M_{rot}^{j} | \pi_{home}) \times P(\lambda_{home}^{j} | M_{rot}^{j} H_{dist}^{j} H_{dir}^{j} \pi_{home}) \\ \times P(M_{trans}^{j} | \pi_{home}) \times P(\tau_{home}^{t} | M_{trans}^{j} H_{dist}^{j} \pi_{home}) \end{array} \right] \\ P(H_{dir}^{0} H_{dist}^{0} Z H_{dir}^{0} Z H_{dist}^{0} \lambda_{home}^{0} \tau_{home}^{0} M_{rot}^{0} M_{trans}^{0} | \pi_{home}). \end{array} \right. \end{aligned} \tag{5.12}$$

The dynamic models in the first line are the same as in the previous chapter. The sensor models in line 2 are also unchanged. The motor inverse model is here replaced by its coherence-based fusion form. As there are two motor variables (M_{rot} and M_{trans}), it is necessary to have two coherence variables (λ_{home} and τ_{home}, respectively) in each elementary filter. Because it is easy to convert an inverse model into a coherence-based fusion model by applying equation 5.6, only the inverse motor models are presented here.

In the previous chapter, the inverse motor model showed a dependence between the motor commands and all state variables (as defined in equation 4.15). Here, the motor commands are described inside each elementary filter, and this implies that the motor model shows dependencies on the elementary filter subsets of variables, as in the above joint distribution.

The specified robot behaviour has well defined movements in relation to home, prey or predator. For example, the motor model in the prey elementary filter defines the desired robot movement in relation to the prey. Similar reasoning is adopted for the home and predator elementary filters.

The inverse model for rotation speed in the home elementary filter indicates moving in the home direction if home is close (distance smaller than two, using the same knowledge as in the previous chapter). If home is not close, it gives no indication (uniform distribution over rotation speed).

$$\begin{aligned} & P(M_{rot}^{j} | H_{dist}^{j} H_{dir}^{j} \pi_{home}) \\ & = \left\{ \begin{array}{l} \mathcal{G}(M_{rot}^{j}), \mu = H_{dir}^{j}, \sigma_{rot}, \text{if } H_{dist}^{j} < 2; \\ \text{Uniform, otherwise.} \end{array} \right. \end{aligned} \tag{5.13}$$

When the situation indicated by the filter states has no special recommendation about motor commands, a uniform distribution is applied. The same reasoning can be applied when writing the inverse motor model for the prey filter. For the predator filter, we have:

$$
\begin{aligned}
&P(M^j_{rot}|X^j_{dist}X^j_{dir}\pi_{pred}) \\
&\quad = \begin{cases} \mathcal{G}(M^j_{rot}), \mu = X^j_{dir} - 180, \sigma_{rot}, \text{if } X^j_{dist} \leq 3 \text{ and } X^j_{dir} \geq 0; \\ \mathcal{G}(M^j_{rot}), \mu = X^j_{dir} + 180, \sigma_{rot}, \text{if } X^j_{dist} \leq 3 \text{ and } X^j_{dir} < 0; \\ \text{Uniform, otherwise.} \end{cases}
\end{aligned} \tag{5.14}
$$

The prediction, estimation and motor question are now asked of each elementary filter. The prediction and motor question (the ones requiring more computational efforts) are now calculated over a smaller number of state variables (two variables each filter, instead of six in the previous chapter). The prediction and estimation questions are employed recursively in each elementary filter, and there is no global recursion. Consequently, a global filter answer is essential only for the motor question.

The prediction question for the home filter is as follows:.

$$P(H^t_{dir}H^t_{dist}|zh^{0:t-1}_{dir}zh^{0:t-1}_{dist}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t-1}_{home}\tau^{0:t-1}_{home}\pi_{home}).$$

For the estimation question, present-time observations are added on the right side:

$$P(H^t_{dir}H^t_{dist}|zh^{0:t}_{dir}zh^{0:t}_{dist}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t-1}_{home}\tau^{0:t-1}_{home}\pi_{home}).$$

The motor question for the home filter is:

$$P(M^t_{rot}M^t_{trans}|zh^{0:t}_{dir}zh^{0:t}_{dist}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t}_{home}\tau^{0:t}_{home}\pi_{home}).$$

Motor question answers from the elementary filters are combined into a global distribution, from which motor values are drawn. Defining the set of all observation variables as:

$$z^t_{lf} = zh^t_{dir}zh^t_{dist}zx^t_{dir}zx^t_{dist}zy^t_{dir}zy^t_{dist}, \tag{5.15}$$

we find the expression for motor global distribution:

$$
\begin{aligned}
&P(M^t_{rot}M^t_{trans}|z^{0:t}_{lf}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t}_{home}\tau^{0:t}_{home}\pi) \\
&\propto \left| \begin{array}{l} P(M^t_{rot}M^t_{trans}|zh^{0:t}_{dir}zh^{0:t}_{dist}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t}_{home}\tau^{0:t}_{home}\pi_{home}) \\ \times P(M^t_{rot}M^t_{trans}|zx^{0:t}_{dir}zx^{0:t}_{dist}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t}_{home}\tau^{0:t}_{home}\pi_{pred}) \\ \times P(M^t_{rot}M^t_{trans}|zy^{0:t}_{dir}zy^{0:t}_{dist}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t}_{home}\tau^{0:t}_{home}\pi_{prey}). \end{array}
\end{aligned} \tag{5.16}
$$

Simulation Results

Figure 5.2 shows a sequence of snapshots of a simulation run, with the same initial conditions as employed in the previous chapter (Figure 4.6). The general behaviour of the robot is unchanged: it escapes from the predator, opportunistically captures a prey, and goes towards home. Unlike the filter in the previous chapter, here the robot actually stops at home.

To analyse the fusion of motor commands, curves for rotation speed are shown at the right of each snapshot in Figure 5.2. At time $t = 0$, seen in Figure 5.2(a), the robot is far away from prey and home: the motor command propositions from the prey and home filter are nearly uniform. However, the predator is dangerously near the robot, and the motor command proposition from the predator filter is then a clear indication to go forward. The command resulting from fusion is almost equal to the predator proposition.

The robot moves straight away from the predator, and Figure 5.2(b) shows time $t = 8$: it is now far from the predator, and the motor proposition from predator filter is less relevant. There is now a prey very near the robot, and the motor command proposition from the prey filter indicates the robot should turn right to follow the prey direction. As before, home is considered to be too far away, and the motor command proposition from the home filter is still near uniform. Fusion of these propositions results in the robot following the prey direction.

After chasing and capturing the prey, the robot changes its direction to escape from the approaching predator. At time $t = 28$, shown in Figure 5.2(c), the robot is going towards home. The predator is not far away, and the predator filter proposes a direction to follow to escape. Home is near, and the home filter proposes its direction as the motor command, which does not conflict with the direction proposed by the predator filter. A new prey has appeared in the corner, and even though it is not close, the prey filter proposes that the robot should follow its direction. The resulting fusion is the home direction, and following this command, the robot arrives home.

The results shown above illustrate utilisation of coherence-based fusion in the same way as sensor fusion. They also show that defining motor commands separately for each filter and combining then afterwards can lead to reasonable results.

5.1.4 Additional Main Assumptions

- State and observation variables can be assembled in mutually exclusive subsets.

  Conditional on the dynamic model, state variables can be grouped into mutually exclusive subsets. Observation variables can also be grouped into disjoint subsets, conditional on independence in the observation model, and each observation subset is related to one and only one of the state subsets.

- An inverse model for motor commands is more convenient to write than a direct model.

  Even if state and observation variables are grouped in mutually exclusive subsets, motor commands are still common variables to all subsets. Using coherence-based fusion, it is possible to have independent subsets of coherence variables relating states to motor commands. Motor inverse models are then easily rewritten using coherence variables, and the fusion of subset proposals for motor commands is possible and straightforward.

5.2 Requirements Analysis

5.2.1 Space Complexity

Utilisation of elementary filters is equivalent to utilisation of a global filter; therefore, only elementary filter joint distribution terms must be kept in memory, together with answers to the questions asked in each filter.

It is not necessary to keep global predictions and estimations in memory, as they are not reused: as discussed in Section 5.1.2, the recursive loop occurs inside each elementary filter.

Furthermore, only a half table ($\lambda_i^j = 1$) is strictly necessary for storing the coherence-based model for motor commands. Consistently with the previous choice of not considering particular simplifications, the whole table is considered in complexity analysis wherever coherence- based models are applied.

For each elementary filter, the following tables are necessary.

- A priori information about states $P(S_i^{j-1}|\pi_i)$ requires $n^{N_{s_i}}$ cells.
- The dynamic model $P(S_i^j|S_i^{j-1}M^{j-1}\pi_i)$ requires $n^{N_{s_i}+N_{s_i}+N_m}$ cells.
- The sensor model $P(Z_i^j|S_i^j\pi_i)$ requires $n^{N_{z_i}+N_{s_i}}$ cells.
- The motor model in coherence-based fusion form $P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i)$ requires $2n^{N_{s_i}+2N_m}$ cells.
- Prediction and Estimation questions in elementary filters are both stored, requiring $2n^{N_{s_i}}$ cells.
- The motor question in the elementary filters requires n^{N_m} cells.

Additionally, it is necessary to store the global distribution for the motor question, requiring n^{N_m} cells. Summing all necessary cells to store the above terms we have:

$$\mathcal{M} = (N_i + 1)n^{N_m} + \sum_{i=1}^{N_i} \left[\begin{array}{c} n^{2N_{s_i}+N_m} + n^{N_{z_i}+N_{s_i}} \\ +2n^{N_{s_i}+2N_m} + 3n^{N_{s_i}} \end{array} \right]. \tag{5.17}$$

In our illustrative example, 111 844 probability values must be stored, in contrast with 1.2×10^{10} for Example 2 in the previous chapter. The reduction of the necessary amount of memory occurs in all terms in the joint distribution where states are present.

The amount of memory required to store the Bayes filters proposed in this chapter thus has an upper bound of the following function:

$$\mathcal{M} = O\left(\begin{array}{c} n^{2max(N_{s_i})+N_m} + n^{max(N_{z_i}+N_{s_i})} \\ +n^{max(N_{s_i})+2N_m} + N_i \times n^{N_m} \end{array} \right). \tag{5.18}$$

Comparing with the space complexity of the previous chapter in equation 4.30:

$$\mathcal{M} = O(n^{2N_s+N_m} + n^{N_s+N_z} + n^{N_s+2N_m}), \tag{5.19}$$

we can see that the space complexity now increases exponentially with the number of state variables in the biggest subset and linearly with the number of elementary filters. The green term in both equations above is related to the dynamic model, and the brown terms are related to the motor model. In equation 5.18, the exponent related to the number of state variables now concerns only the biggest subset of elementary filters.

The sizes of observation variable subsets are also relevant, as shown in the blue term in both equations above. This influence can be reduced by applying sensor model schema whenever possible, as presented in Chapter 3. The number of probability values changes to:

$$\mathcal{M} = (N_i + 1)n^{N_m} + \sum_{i=1}^{N_i} \left[\begin{array}{l} n^{2N_{s_i}+N_m} + N_{z_i} n^{N_{s_i}+1} \\ +2n^{2N_m+N_{s_i}} + 3n^{N_{s_i}} \end{array} \right]. \tag{5.20}$$

The time complexity loses its exponential dependence on N_{z_i}, as in:

$$\mathcal{M} = O\left(\begin{array}{l} n^{2max(N_{s_i})+N_m} + max(N_{z_i}) n^{max(N_{s_i})} \\ +n^{2N_m+max(N_{s_i})} + N_i \times n^{N_m} \end{array} \right). \tag{5.21}$$

In the previous chapter, the complexity under sensor fusion was given by equation 4.31:

$$\mathcal{M} = O(n^{2N_s+N_m} + N_z n^{N_s} + n^{N_s+2N_m}). \tag{5.22}$$

It can be seen that for the observation variables also, it is possible to reduce the exponential dependence to a linear one, related to the biggest subset of observation variables.

In robotic applications, the number of observation variables is often bigger than the number of state variables: the reduction of exponential to linear dependence by applying sensor fusion becomes an efficient way to reduce space complexity in these situations.

It is also reasonable to consider that the number of elementary filters is considerably smaller than the number of cases of state variables in the biggest subset. Mathematically:

$$N_i \ll n^{max(N_{s_i})}. \tag{5.23}$$

Consequently, a term implying dependence on the number of elementary filters is not relevant in space complexity, and we can conclude that the required memory complexity does not depend on the number of elementary filters.

5.2.2 Time Complexity

In this model, calculations are done in the following order.

1. Prediction in each elementary filter according to equation 5.7 (and repeated below) requires $4n^{2N_{s_i}} - n^{N_{s_i}}$ operations.

$$\begin{array}{l} P(S_i^t|m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}\pi_i) \\ \quad \propto \ \sum_{S_i^{t-1}} \left[\begin{array}{l} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\ \times P(m^{t-1}|\pi_i) \times P(\lambda_i^{t-1}|m^{t-1}S_i^{t-1}m^{t-2}\pi_i) \\ \times P(S_i^{t-1}|m^{0:t-2}z_i^{0:t-1}\lambda_i^{0:t-2}\pi_i) \end{array} \right]. \end{array}$$

2. Estimation in each elementary filter (equation 5.8 and shown below) requires $n^{N_{s_i}}$ operations.

$$\begin{array}{l} P(S_i^t|m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}\pi_i) \\ \quad \propto \left| \begin{array}{l} P(z_i^t|S_i^t\pi_i) \\ \times P(S_i^t|m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}\pi_i). \end{array} \right. \end{array}$$

3. The motor command in each elementary filter is given by equation 5.10 (reproduced below) requires $3n^{N_{s_i}+N_m} - n^{N_m}$ operations.

$$\begin{array}{l} P(M^t|m^{0:t-1}z_i^{0:t}\lambda_i^{0:t}\pi_i) \\ \quad \propto \ \sum_{S_i^t} \left[\begin{array}{l} P(M^t|\pi_i) \times P(\lambda_i^t|M^tS_i^tm^{t-1}\pi_i) \\ \times P(S_i^t|m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}\pi_i) \end{array} \right] . \end{array}$$

4. The global motor command, described in 5.11 (repeated below) requires $N_i n^{N_m}$ operations.

$$\begin{array}{l} P(M^t|m^{0:t-1}z^{0:t}\lambda^{0:t}\pi_{lf}) \\ \quad \propto \ \prod_{i=1}^{N_i} P(M^t|m^{0:t-1}z_i^{0:t}\lambda_i^{0:t}\pi_i). \end{array}$$

Summing up the computations to solve all questions, we have:

$$\mathcal{C} = \sum_{i=1}^{N_i} \{4n^{2N_{s_i}} + 3n^{N_{s_i}+N_m}\}. \tag{5.24}$$

Considering our illustrative example, only 11 256 operations are now necessary at each time step. This drastic reduction compared with the previous chapter, where $1.4 \times$

10^8 operations were required, is the result of the conditional independence between the subsets of state and observation variables.

The number of operations (sums or multiplications) necessary for solving the ultimate question now has the upper bound:

$$\mathcal{C} = O(n^{2max(N_{s_i})} + n^{max(N_{s_i})+N_m}). \tag{5.25}$$

In the previous chapter, the time complexity was given by equation 4.33:

$$\mathcal{C} = O(n^{2N_s} + n^{N_m+N_s}). \tag{5.26}$$

In both these equations, the green term is related to prediction: the double rate in the exponential is present in both terms, but in equation 5.25, the size of the biggest state subset instead of the number of state variables is used. The size of the biggest state subset is also present in the brown term, which is related to asking the motor question. Consequently, the size of the biggest subset of state variables determines how the calculation time scales with system size.

Comparing the above time complexity expressions, it is easy to remark that the size of the state subsets (and not the number of subsets) is responsible for the decreasing complexity.

As before, the number of observation variables becomes relevant for the time complexity when sensor fusion is employed. It alters the effort necessary to calculate the estimation question, and the number of operations to be executed at each time step becomes:

$$\mathcal{C} = \sum_{i=1}^{N_i} \{3n^{2N_{s_i}} + (N_{z_i} - 1)n^{N_{s_i}} + 2n^{N_{s_i}+N_m}\}. \tag{5.27}$$

Consequently, time complexity is defined by:

$$\mathcal{C} = O(n^{2max(N_{s_i})} + max(N_{z_i})n^{max(N_{s_i})} + n^{max(N_{s_i})+N_m}). \tag{5.28}$$

Comparing with equation 4.35 from the previous chapter:

$$\mathcal{C} = O(n^{2N_s} + N_z n^{N_s} + n^{N_m+N_s}), \tag{5.29}$$

we can see that the biggest subset of observation variables influences time complexity linearly, and applying sensor fusion actually increases time complexity. Because of the reduction in space complexity, the decision to employ sensor fusion schema results from a compromise between the increase in time complexity, space complexity reduction and application-specific characteristics.

5.2.3 Programming

Compared with the Bayes filter proposed in the previous chapter, programming becomes easier inside each elementary filter. The independence condition between state

and observation variables is enough to ensure that programming dynamic and sensor models in each elementary filter can be isolated from all other elementary filters.

Even if motor command variables are common to all elementary filters, motor models are written individually for each elementary filter. Knowledge about other elementary filters is not required, and whenever variable subset values have no relevant information about motor commands, the uniform distribution is applied.

The main difficulty in programming this version is the choice of variables to fulfil the conditions of mutually independent subsets in relation to the dynamic and sensor models. In our example, we have chosen state variables to represent the egocentric position [3] of objects or features in the world, and this has resulted in one independent subset of state variables for each relevant feature.

An interesting hypothesis to be examined is whether such a representation is a sufficient condition for having disjoint subsets and the kind of application for which this is appropriate.

Learning

Utilisation of elementary filters can reduce complexity when learning dynamic and observation models: it is possible to work on each elementary filter separately. Even if it is also possible to learn motor commands separately, it is necessary to keep in mind that because of the later fusion of motor commands, uniform representations should be favoured when the elementary filter state subsets have no relevant proposal for motor commands.

5.3 Discussion

5.3.1 Robotics

Structured State Space

The reduction of time complexity when employing Bayesian modelling approaches is the subject of several research projects. For POMDP, for example, two main issues are the exponential complexity in relation to state spaces and the time horizon for calculating policy. As our solution does not plan over a time horizon, we are mostly interested in reducing the influence of state space over on-line calculations so that they can be executed in a robotic real-time task.

[3] We consider that the position of an object in the environment can be represented using an egocentric or an allocentric frame of reference, which are coordinate systems in which positions can be specified in two dimensions. An egocentric frame of reference is defined by the robot body: left to right and front to back axis, for example. An allocentric frame is composed of orthogonal axes set outside the robot. For more details, see Bryant [1992].

Solutions proposed in the literature for computing state-related questions (called *computation of belief state*) are mostly compact representations of the state space (as discussed by Boyen [2002]) or approximate inference algorithms, described in Thrun [2002a].

The factored Markov decision process (factored MDP) and its counterpart factored partially observable Markov decision process (factored POMDP) both exploit problem-specific structure to represent large and complex dynamic systems compactly [Guestrin, 2003; Poupart, 2005].

The main idea of factored representation in MDP is to split the state space into subsets, disjointed or weakly coupled in relation-state dynamics. Then, algorithms for computing control policy are developed that cope with the state space separation [Parr, 1998; Lane and Kaelbling, 2001; Guestrin, 2003].

In POMDP, factored representations find additional problems in the observation and action spaces: conditional independence or even context-specific independence allows us to write dynamic and observation models in factored form [Poupart, 2005, Chapter 4]. Williams et al. [2005] proposed a factored POMDP for a dialogue management application.

Our proposal is an attempt to reduce this complexity by imposing some constraints on dependencies in the dynamic and observation models. These constraints must be verified in the early stages of conception of the robot task model.

Our approach is very similar to the compact representation proposed by Boyen [2002], which groups state variable subsets with weak dependencies with other subsets in a cluster. The compact representation is then built using a cluster-structured filter. It is shown that the accumulated error between this compact representation utilisation and the complete system diminishes with time according to the choice of clusters: a technique called *Contraction Analysis* is proposed to analyse different possible choices of clusters so that a good complexity reduction is achieved and the error is made reasonable. Another technique called *Projection Analysis* is proposed to analyse whether removed dependencies, even if weak, could affect a posteriori estimation in a longer time.

We argue that the choice of state and observation variables is decisive for defining dependencies in dynamic and observation models. More specifically, referential choice is crucial when dealing with dependencies. In the illustrative example, we have shown state and observation variables fulfilling these conditions. These variables were defined as being egocentric in relation to the robot: the distance of the predator is defined as the distance from predator to robot. As predator, prey and home are environment features having no mutual influence, variable subsets fulfilling the constraints are easily defined.

5.3.2 Cognition

The cognitive interpretations of the mathematical hypotheses added in this chapter (see decomposition in equation 5.1) are the following:

- The Central Nervous System is structured in modules where the dynamic mod-

els of the evolution of internal states and the corresponding sensor models are independent of the other modules:

$$\prod_{i=1}^{N_i} \left[P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \right]$$
$$\prod_{i=1}^{N_i} \left[P(Z_i^j | S_i^j \pi_i) \right].$$

- Each of these modules has an opinion on the adequate motor commands and their diverging opinions can be merged in a unique coherent one:

$$P(M^j | \pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \right].$$

The different steps to answer the ultimate question are:

- Predict the internal state corresponding to each area of interest (equation 5.7).
- Update the internal states of each area of interest, knowing observations specific to each area of interest (equation 5.8).
- Ask each area of interest for its motor recommendation (equation 5.10).
- Finally, merge their opinion in a coherent whole to decide of the motor commands (equation 5.11).

States prediction and estimation now takes place locally in each module. It can be demonstrated mathematically that the ultimate question may be solved by local evaluations of motor commands in each module and later assembled by a simple product before making the motor decision, as in equation 5.11.

Local computation is responsible for the reduction of exponent in time and space complexity: instead of depending on the size of state space, this exponent depends now on the size of the biggest local state sub-space (equations 5.18 and 5.28). This reduction is obtained because the necessary summations, for prediction and motor command questions, have only to be done on the local state sub-space (as may be seen by the S_i^{t-1} and S_i^t indexes in equations 5.7 and 5.10 respectively). This is the third important step to keep the computation tractable.

We propose to call these modules "areas of interest" as they collect in a coherent whole all the sensory information necessary to build a synthetic internal representation of a subject in order to decide the adequate motor commands. Examples of these areas of interest are environment features relevant to the behaviour: predators, preys and home.

5.4 Summary and Open Questions

In this chapter, we proposed conditions for reducing the complexity of a Bayes filter for answering the ultimate question. These conditions are related to conditional independence between subsets of state and observation variables. Furthermore, the definition of the motor command model is made using coherence-based fusion. Consequently, each disjoint subset defines another Bayes filter, called an elementary filter, with the joint distribution:

$$
\begin{aligned}
& P(M^{0:t} S_i^{0:t} Z_i^{0:t} \lambda_i^{0:t} | \pi_i) \\
& = \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i). \end{array} \right.
\end{aligned}
$$

Questions now are asked of each elementary filter, and whenever a summation over state space is required, it will concern only the subset of state variables for this specific filter.

The prediction question in each elementary filter is then:

$$
\begin{aligned}
& P(S_i^t | m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} \pi_i) \\
& \propto \sum_{S_i^{t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(m^{t-1} | \pi_i) \times P(\lambda_i^{t-1} | m^{t-1} S_i^{t-1} m^{t-2} \pi_i) \\ \times P(S_i^{t-1} | m^{0:t-2} z_i^{0:t-1} \lambda_i^{0:t-2} \pi_i) \end{array} \right].
\end{aligned}
$$

The estimation question uses the prediction expression:

$$
\begin{aligned}
& P(S_i^t | m^{0:t-1} z_i^{0:t} \lambda_i^{0:t-1} \pi_i) \\
& \propto \left| \begin{array}{l} P(z_i^t | S_i^t \pi_i) \\ \times P(S_i^t | m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} \pi_i). \end{array} \right.
\end{aligned}
$$

The motor question in each elementary filter depends on the estimation:

$$
\begin{aligned}
& P(M^t | m^{0:t-1} z_i^{0:t} \lambda_i^{0:t} \pi_i) \\
& \propto \sum_{S_i^t} \left[\begin{array}{l} P(M^t | \pi_i) \times P(\lambda_i^t | M^t S_i^t m^{t-1} \pi_i) \\ \times P(S_i^t | m^{0:t-1} z_i^{0:t} \lambda_i^{0:t-1} \pi_i) \end{array} \right].
\end{aligned}
$$

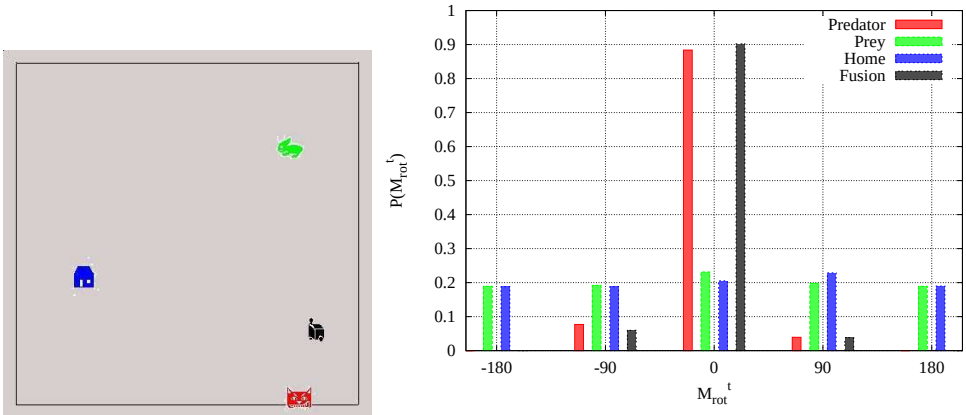
Before drawing values for the motor commands, it is necessary to combine the answers from all the elementary filters:

$$
\begin{aligned}
& P(M^t | m^{0:t-1} z^{0:t} \lambda^{0:t} \pi_{lf}) \\
& \propto \prod_{i=1}^{N_i} P(M^t | m^{0:t-1} z_i^{0:t} \lambda_i^{0:t} \pi_i).
\end{aligned}
$$

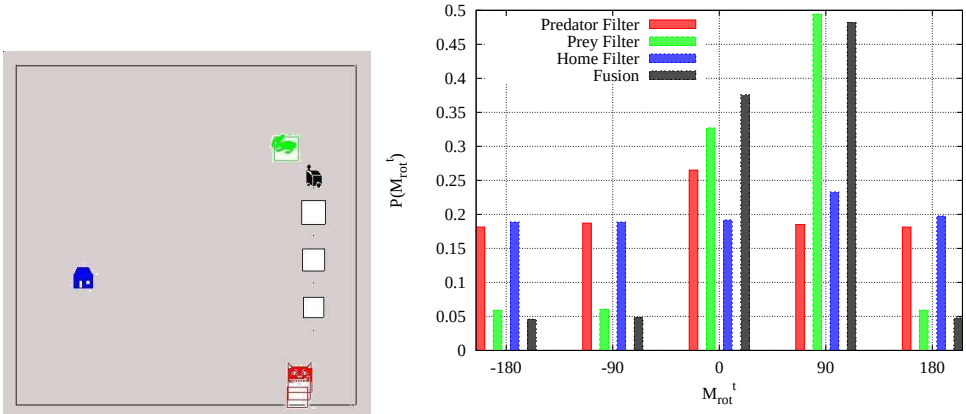
Conditional independence can then be fully exploited with the help of coherence-based fusion.

One of the benefits of the proposal in this chapter is that writing motor models becomes easier, as it is possible to write several smaller models instead of a unique big one dealing with many variables. However, writing motor models is now limited to simple behaviours, when it is still possible to relate states directly to motor commands.

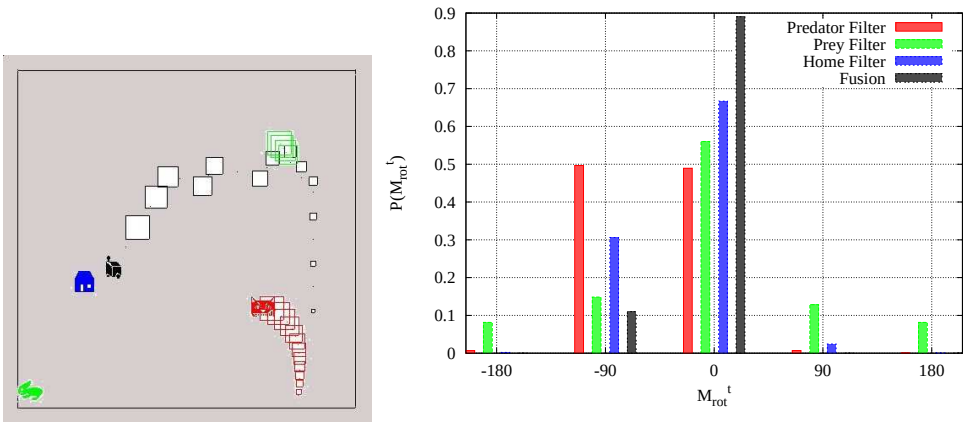
In the next chapter, we propose a method to include an action selection mechanism in the Bayes filter to cope with more complex specified behaviours.



(a) $t = 0$



(b) $t = 8$



(c) $t = 28$

Figure 5.2: Sequence of snapshots for simulation of Example 2. On the left, simulation snapshots and, on the right, curves for rotation speed distributions, one for each filter answer and the resulting fusion.

Chapter 6

Behaviour Coordination

Key Points:

- **Motor commands form a repertoire of simple and independent motor patterns.**
- **A behaviour variable switches between motor patterns.**
- **Behaviour choice is the result of the fusion of Elementary filter proposals.**

The utilisation of elementary filters proposed in the previous chapter helps to reduce complexity in space and time, and it also simplifies writing motor models.

Motor commands can be grouped in collections of simple and independent motor patterns: follow an object, go straight ahead, and so on. Each pattern is called here a basic behaviour, and in this chapter, we propose the addition of a behaviour selection mechanism to the Bayes filter of the previous chapter.

Selection of the most pertinent behaviour is made employing fusion of elementary filter proposals, which are the result of a local strategy of behaviour choice in each elementary filter according to its own subset of variables.

6.1 Bayesian Program

When programming a robot to execute a specified behaviour, combinations of simple motor command patterns are often sufficient for creating more complex movements. A simple and straightforward example of a motor pattern is the combination of wheels speed in order to make the robot go straight ahead (both wheels turn in the same direction) or turn right or left (wheels turn in opposite directions). Another example is when the robot turns the head towards one direction and the body follows the head.

The filter of the previous chapter allowed motor commands proposed by each elementary filter, and the result of combining these proposals was applied to actuators. However, the proposals from the elementary filters can conflict or be inconsistent between themselves. For example, the head going towards one direction and the body following another one.

In this chapter, motor commands are grouped in patterns, and each elementary filter rates every possible motor pattern as more or less adequate considering its own subset of state and observation variables. Fusion of elementary filter proposals allows a decision about which motor command pattern is more pertinent and allows its coherent execution with the assistance of all elementary filter motor models.

6.1.1 Description

Additional Relevant Variables

To coordinate the execution of these patterns better, we propose to add a set of behaviour variables, denoted B^t, common to all elementary filters. The semantics of these variables are related to the patterns found in the motor commands: each variable corresponds to a group of motor patterns, one value for each pattern. One motor pattern corresponds to a set of motor commands applied simultaneously on motor actuators and it can depend on state variables.

Behaviour variables are common to all elementary filters, and coherence-based fusion is applied to describe the behaviour selection model: β^t stands for the set of coherence variables associated with the behaviour variables B^t. The number of coherence variables depends on the number of behaviour variables and the number of elementary filters: one variable β is necessary for each variable B in each elementary filter.

At this point, it is important to make a clear distinction between the robot's specified behaviour, the resulting behaviour and the behaviour variables described above.

- The *robot's specified behaviour* is a description of the desired robot task: it can be achieved or not.

- The *robot's resulting behaviour* is the result of executing the controller in a physical or simulated robot. An ideal robot controller would make the resulting behaviour equal to the specified behaviour.

- The *behaviour variables* introduced here are directly related to the distinctive and coordinated motor commands: when behaviour variables are assigned specific values, certain motor commands are executed. Execution of these motor commands in a coordinated way generates the robot's resulting behaviour.

Decomposition

The joint distribution for a general elementary filter is as shown below.

$$P(M^{0:t} S_i^{0:t} Z_i^{0:t} B^{0:t} \lambda_i^{0:t} \beta_i^{0:t} \pi_i)$$

$$= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i). \end{array} \right. \tag{6.1}$$

Inside the time product, the first line consists of the dynamic model, and it is not changed from the dynamic model in the previous chapters. The second line contains the sensor model, also unchanged.

The third line shows the behaviour model, in the coherence-based fusion form:

$$P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i).$$

Behaviour at the present time depends on the behaviour selected at the previous time and on present time states.

The relation between selected behaviour at the previous and present time aims to ensure continuity (called persistence) in the execution of a behaviour, even if states show a reduction of environment stimuli. Dependence on the present state means favouring reactivity to environment changes.

It is easier to define the behaviour selection model initially as an inverse model and later transform it to the coherence-based fusion form, which ensures the fusion of elementary proposals. The conversion of a behaviour selection inverse model into a coherence-based fusion model follows the expressions below.

$$\begin{array}{l} P([\beta_i^j = 1] | B^j S_i^j B^{j-1} \pi_i) = P(B^j | S_i^j B^{j-1} \pi_i), \\ P([\beta_i^j = 0] | B^j S_i^j B^{j-1} \pi_i) = 1 - P(B^j | S_i^j B^{j-1} \pi_i). \end{array} \tag{6.2}$$

The motor command model,

$$P(\lambda_i^t | M^t S_i^t B^t M^{t-1} \pi_i),$$

in the fourth line of the joint distribution shown in equation 6.1, now includes the dependence on the behaviour variable.

It is interesting to emphasise the role of the behaviour variable in the motor model. Behaviour variables take control of switching between motor models: each value corresponds to a motor pattern defined by a specific motor model. Considering only one behaviour variable with n_b possible values, we have the following motor model:

$$\begin{array}{l} P(M^t | S_i^t B^t M^{t-1} \pi_i) \\ \quad = \left\{ \begin{array}{l} P(M^t | S_i^t M^{t-1} [B^t = b_1] \pi_i); \\ P(M^t | S_i^t M^{t-1} [B^t = b_2] \pi_i); \\ \ldots \\ P(M^t | S_i^t M^{t-1} [B^t = b_{n_b}] \pi_i). \end{array} \right. \end{array} \tag{6.3}$$

Each term in the above equation corresponds to one motor pattern:

- $P(M^t|S_i^t M^{t-1}[B^t = b_1]\pi_i)$ is the motor model of behaviour b_1;
- $P(M^t|S_i^t M^{t-1}[B^t = b_2]\pi_i)$ is the motor model of behaviour b_2, and so on.

The last line of the joint distribution shows the initial conditions of all relevant variables.

Bayesian network representations of both elementary and global filter joint distributions are presented in appendix C, Figures C.1 and C.2, respectively.

6.1.2 Utilisation

Compared with utilisation in the previous chapter, one more question is asked of each elementary filter, called the behaviour selection question. Demonstration of the recursive calculation and the relations between elementary filter questions and the global filter questions are thoroughly described in Appendix C.

Bayes filter utilisation is summarised in Figure 6.1. Fusion of elementary filter answers is executed twice: first for behaviour selection, then for motor commands. State variable estimation is done after behaviour selection, which ensures the recursive calculation for the next time step.

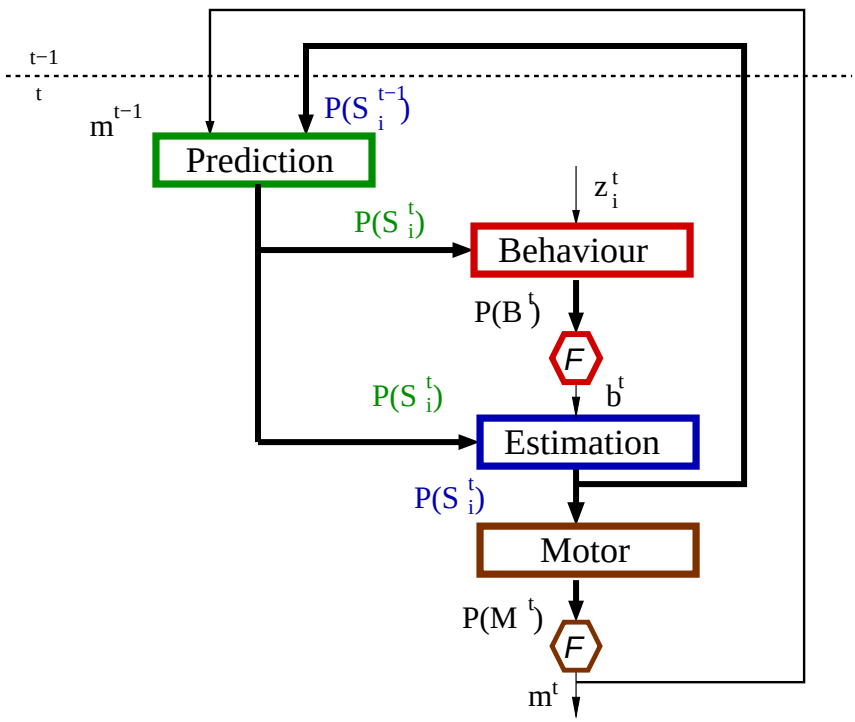


Figure 6.1: Utilisation of Bayes filter with behaviour coordination.

Prediction

$$
\begin{aligned}
& P(S_i^t | z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i) \\
& \propto \sum_{S_i^{t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(m^{t-1} | \pi_i) \times P(\lambda_i^{t-1} | m^{t-1} S_i^{t-1} b^{t-1} m^{t-2} \pi_i) \\ \times P(S^{t-1} | z_i^{0:t-1} b^{0:t-1} m^{0:t-2} \lambda_i^{0:t-2} \beta_i^{0:t-1} \pi_i) \end{array} \right] .
\end{aligned} \tag{6.4}
$$

The answer to the prediction question has not changed from the previous chapter: the expression includes the dynamic model, the influence of motor command choice and estimation at the previous time. States are not known, so prediction results from a weighted sum over the state variables at the previous time.

The Behaviour Selection Question in an Elementary Filter

The question about behaviour selection is asked to each elementary filter joint distribution, and answers from all elementary filters are combined before taking a decision.

The behaviour selection question for a general elementary filter is then:

$$
\begin{aligned}
& P(B^t | z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \pi_i) \\
& \propto \sum_{S_i^t} \left[\begin{array}{l} P(z_i^t | S_i^t \pi_i) \\ \times P(B^t | \pi_i) \times P(\beta_i^t | B^t b^{t-1} S_i^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i) \end{array} \right] .
\end{aligned} \tag{6.5}
$$

The left part of the question consists of the set of behaviour variables, and the right part has all known values from the previous time as well as observations from the present time instant. The coherence variable for behaviour variable β_i^t is assigned a value equal to one.

The sensor model in the first line updates the prediction over states with present observations. The second line contains the behaviour selection model, in the form of coherence-based fusion. As β_i^t is equal to one, only coherent relations between the present-time behaviour and the present- time states are taken into account.

The last line consists of the prediction expression calculated in equation 6.4. The behaviour question is weighted over the probabilities of the state variables at the present time because their values are not known.

The Behaviour Selection Question in Global Filter

When answers to the behaviour selection question from all elementary filters are available, they are combined to produce the global expression for behaviour selection. According to the demonstration in Appendix C, asking the behaviour selection question of a global filter is proportional to the product of the elementary filter answers to the behaviour question, as shown in the equation in equation 6.6 below.

$$P(B^t|z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\pi_{bs}) \propto \prod_{i=1}^{N_i} P(B^t|z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\pi_i). \tag{6.6}$$

Values for the behaviour variables at the present time b^t are drawn according to the global probability distribution calculated above.

Estimation

Estimation is now calculated considering observations at the present time as well as the values drawn for the behaviour variables.

$$P(S_i^t|z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\pi_i) \propto \left| \begin{array}{l} P(z_i^t|S_i^t\pi_i) \\ \times P(b^t|\pi_i) \times P(\beta_i^t|b^t b^{t-1} S_i^t \pi_i) \\ \times P(S_i^t|z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\pi_i). \end{array} \right. \tag{6.7}$$

This expression shows the estimation answer in each elementary filter. In the first line, the sensor model updates the prediction with observations. The second line introduces the behaviour model's influence over the states: because the values for the present behaviour were already drawn, the state variables are updated in relation to these values.

The sensor model and behaviour model work together in this expression as in the sensor fusion schema shown in Chapter 3. Actually, both provide evidence on state variables based on known values: observation values for the sensor model, and the behaviour value in the behaviour model.

The Motor Question in the Elementary Filter

Because the present behaviour has already been selected, the variable b^t now appears in the right part of the motor command question for each elementary filter:

$$P(M^t|z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t}\beta_i^{0:t}\pi_i) \propto \sum_{S_i^t} \left[\begin{array}{l} P(M^t|\pi_i) \times P(\lambda_i^t|M^t S_i^t b^t m^{t-1}\pi_i) \\ \times P(S_i^t|z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\pi_i) \end{array} \right]. \tag{6.8}$$

This expression is similar to the motor command question for the Bayes Filter described in the previous chapter: the first line contains the command model in the form of coherence-based fusion, and the second line consists of estimation of the state variables. As before, the present-time state variable values are not known, and to solve this expression, it is necessary to weight the motor command model by the state estimations.

The Motor Question in the Global Filter

Analogously to the previous chapter, elementary filter answers are gathered together into a global distribution for the motor command variables.

$$P(M^t|z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\pi_{bs}) \propto \prod_{i=1}^{N_i} P(M^t|z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t}\beta_i^{0:t}\pi_i). \tag{6.9}$$

The motor commands values are drawn from the global distribution calculated above, and the drawn values are applied to the robot motor controls.

6.1.3 Illustration Example

We illustrate the utilisation of the behaviour selection model with the same example as in the previous chapter, except that the behaviour specification is more complex here.

In addition to escaping the predator, chasing the prey, and going home as described in Example 2, Chapter 4, the robot can stay motionless when it detects a predator, in the hope that the predator does not see it.

Bayesian Program

Motor patterns appearing in the specified robot behaviour are *Escape*, *Go towards prey*, *Go towards home*, and *Stay motionless*. The *Escape* pattern is always related to the predator, while the *Go towards* pattern can be applied to chasing prey and to going home.

These motor patterns concern the control of robot body movements and cannot be executed simultaneously. We have then only one behaviour variable, B, which when assigned a value indicates which motor pattern applies. This behaviour variable is added to the filter of the previous chapter, which can take four values: escape from predator (Escape), chase the prey (Chase), go home (Homing), and stay motionless (Motionless).

Behaviour selection is considered to depend on the behaviour selected at the previous time step and on the distance state variables. The behaviour selection model is written using coherence-based fusion, and one coherence variable β in each filter is associated with the behaviour variable B.

The predator filter, for example, has the following joint distribution:

$$P(X_{dir}^{0:t}X_{dist}^{0:t}ZX_{dir}^{0:t}ZX_{dist}^{0:t}M_{rot}^{0:t}M_{trans}^{0:t}B^{0:t}\lambda_{pred}^{0:t}\tau_{pred}^{0:t}\beta_{pred}^{0:t}|\pi_{pred})$$

$$
= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(X^j_{dist}|X^{j-1}_{dist}\pi_{pred}) \times P(X^j_{dir}|X^{j-1}_{dir}M^{j-1}_{rot}\pi_{pred}) \\ \times P(ZX^j_{dir}|X^j_{dir}\pi_{pred}) \times P(ZX^j_{dist}|X^j_{dist}\pi_{pred}) \\ \times P(B^j|\pi_{pred}) \times P(\beta^j_{pred}|B^jB^{j-1}X^j_{dist}\pi_{pred}) \\ \times P(M^j_{rot}|\pi_{pred}) \times P(\lambda^j_{pred}|M^j_{rot}B^jX^j_{dist}X^j_{dir}\pi_{pred}) \\ \times P(M^j_{trans}|\pi_{pred}) \times P(\tau^j_{pred}|M^j_{trans}B^jX^j_{dist}\pi_{pred}) \end{array} \right] \\ \times P(X^0_{dir}X^0_{dist}ZX^0_{dir}ZX^0_{dist}\lambda^0_{pred}\tau^0_{pred}M^0_{rot}M^0_{trans}B^0\beta^0_{pred}|\pi_{pred}). \end{array} \right. \tag{6.10}
$$

Inside the time product, we can see the dynamic model in the first line, unchanged from the previous chapter, and the sensor model in the second line, also unchanged. The third line shows the behaviour selection model in coherence-based fusion form. The last two lines inside the time product consist of motor models, with the addition of behaviour variable dependence. Outside the time product are the initial conditions for all relevant variables.

Behaviour selection depends on the behaviour selected in the previous time step and also on the predator distance at the present time:

$$
P(\beta^j_{pred}|B^jB^{j-1}X^j_{dist}\pi_{pred}).
$$

Writing the behaviour selection model for the predator filter can be delicate as not all states are available: going home, for example, is one of the strategies for self-defence from the predator, but the predator filter has no knowledge of *home*.

As the behaviour proposals of each elementary filter are combined in fusion, it is more interesting to program behaviour in a proscriptive way: prohibited situations have very low probability values, and recommended behaviours are favoured with higher probability values. If a behaviour is not directly related to a state subset in this elementary filter, it is important to give to the other filters the possibility of choosing it. This is possible using intermediate probability values. A behaviour classified as unwise (not prohibited but not desired) represents a situation that is neither recommended nor forbidden. Only the accumulation of evidence from other filters can make it highly probable.

The behaviour selection model in each elementary filter supplies an expert opinion on each behaviour value, indicating if it is completely forbidden, rather unwise, not relevant, or recommended, in terms of the information available in the elementary filter.

Table 6.1 shows how the behaviour selection inverse model is programmed for the predator filter. If the predator is very close ($X^j_{dist} = 0$), chasing behaviour is forbidden, as it does not offer a solution to escaping from predator attack. Staying motionless is considered to be unwise when the predator is very close. Going home and escaping from the predator are both relevant in this situation, with a higher recommendation for escaping. These relations are used for all values of behaviour selected in the previous time instant and are shown in the first column in Table 6.1.

When the predator is close ($1 \leq X^j_{dist} \leq 2$), the situation is analogous to when the predator is very close, but escaping and going home are equally recommended. This situation appears in the second column in Table 6.1.

X^j_{dist}	0	1,2	3				4			
B^j \ B^{j-1}	*	*	E	C	H	M	E	C	H	M
E	++	++	++	++	++	++	+	∘	∘	∘
C	×	×	×	×	×	×	∘	+	∘	∘
H	+	++	++	++	++	++	∘	∘	+	∘
M	-	-	∘	++	∘	++	∘	∘	∘	+

Table 6.1: Probability table for $P(B^j|B^{j-1}X^j_{dist}\pi_{pred})$, behaviour selection inverse model for predator filter. E: Escape behaviour; C: Chase behaviour; H: Homing behaviour; and M: Motionless behaviour. ++ means highly recommended, + means recommended, ∘ means not relevant, – means unwise and × means forbidden. An asterisk is employed when the situation is the same for all values of one variable. Red cell can be read as *If predator distance is three and robot was chasing it is forbidden to continue chasing.*

If the predator is at a medium distance from the robot ($X^j_{dist} = 3$), what to do next also depends on what the robot was doing, as can be seen in the third to sixth columns in Table 6.1. Escaping and going home are strongly recommended, no matter what the robot was doing, while chase is always forbidden. Motionless is a highly recommended behaviour, unless the robot was already in a defence behaviour such as escaping or homing.

When the predator is far away ($X^j_{dist} = 4$), this filter can only indicate following the previous behaviour, as shown in the rightmost columns in Table 6.1.

Behaviour selection tables for all filters are built in a similar way, and they are shown in Appendix F. Probability values are the same in all tables: ++ is assigned 20, + is assigned 10; ∘ is assigned 1 and – is assigned 10^{-2}; × is assigned 10^{-8}. These values are obviously not normalised, but they are suitable for filling the tables using reference values: highly recommended behaviours are twice as probable as recommended behaviours and 20 times more probable than unconcerned behaviours, and so on. After the table is completed, the probability values are normalised before being used in answering questions.

Returning to the joint distribution 6.10, it can be seen that the motor model has changed compared with the filter in the previous chapter: the behaviour variable is now included in the model dependencies, and it is responsible for switching between motor patterns. When the selected behaviour is not concerned with an elementary filter's state variables, a uniform distribution over the motor commands is employed. When Chase behaviour is selected, for example, the motor command model in the home filter has no relevant suggestion.

When the state variables in the elementary filter are relevant to the execution of the selected behaviour, the motor command model is similar to the one described in the previous chapter. We describe here only the inverse model for rotation speed in the predator filter.

$$
\begin{aligned}
&P(M^j_{rot}|B^j X^j_{dir} X^j_{dist} \pi_{pred}) \\
&= \begin{cases} P(M^j_{rot}|X^j_{dir}[B^t = Escape]\pi_{pred}); \\ P(M^j_{rot}|[B^t = Motionless]\pi_{pred}); \\ P(M^j_{rot}|X^j_{dist} X^j_{dir}[B^t = Chase]\pi_{pred}); \\ P(M^j_{rot}|X^j_{dist} X^j_{dir}[B^t = Homing]\pi_{pred}). \end{cases}
\end{aligned} \tag{6.11}
$$

Behaviours related to the state variables in this filter are Escape and Motionless. If the behaviour is Escape, a dependence also exists on predator direction, and the motor model for rotation speed is analogous to the one in the previous chapter shown in equation 5.14:

$$
\begin{aligned}
&P(M^j_{rot}|[B^t = Escape]X^j_{dir}\pi_{pred}) \\
&= \begin{cases} \mathcal{G}(M^j_{rot}), \mu = X^j_{dir} - 180, \sigma_{rot}, \text{if } X^j_{dir} \geq 0; \\ \mathcal{G}(M^j_{rot}), \mu = X^j_{dir} + 180, \sigma_{rot}, \text{if } X^j_{dir} < 0. \end{cases}
\end{aligned} \tag{6.12}
$$

No state variable is required for behaviour Motionless: its inclusion as relevant in predator filter is mainly because of its classification as a self-defence strategy, related to the predator.

$$
P(M^j_{rot}|[B^t = Motionless]\pi_{pred}) = \mathcal{G}(M^j_{rot}), \mu = 0, \sigma_{rot}. \tag{6.13}
$$

The same expression is used for the two other behaviours: if the predator is far away, a uniform distribution is employed; otherwise, the predator direction is avoided. Distributions are always normalised before utilisation.

$$
\begin{aligned}
&P(M^j_{rot}|X^j_{dist} X^j_{dir}[B^t = Chase]\pi_{pred}) \\
&= \begin{cases} 10^{-12}, \text{if } X^j_{dist} < 4 \text{ and } X^j_{dir} = M^j_{rot}\ ; \\ 1, \text{if } X^j_{dist} < 4 \text{ and } X^j_{dir} \neq M^j_{rot}\ ; \\ \text{Uniform, otherwise.} \end{cases}
\end{aligned} \tag{6.14}
$$

The motor models in the prey and home filters are simple: if the behaviour is Chase (in the prey filter) or Homing (in the home filter), a Gaussian is used, and its mean value is the prey (or home) direction. For other behaviours, a uniform distribution is applied.

The prediction question, as in the previous chapter, is asked to each elementary filter. In the predator filter, for example, we have:

$$
P(X^t_{dir} X^t_{dist}|zx^{0:t-1}_{dir} zx^{0:t-1}_{dist} b^{0:t-1} m^{0:t-1}_{rot} m^{0:t-1}_{trans} \lambda^{0:t-1}_{pred} \tau^{0:t-1}_{pred} \beta^{0:t-1}_{pred} \pi_{pred}).
$$

Following prediction, the behaviour question is asked to each elementary filter. For the predator filter:

$$
P(B^t|zx^{0:t}_{dir} zx^{0:t}_{dist} b^{0:t-1} m^{0:t-1}_{rot} m^{0:t-1}_{trans} \lambda^{0:t-1}_{pred} \tau^{0:t-1}_{pred} \beta^{0:t}_{pred} \pi_{pred}).
$$

Next, the combination of answers from all elementary filters is calculated as:

$$
\begin{array}{l}
P(B^t|z_{lf}^{0:t} b^{0:t-1} m_{rot}^{0:t-1} m_{trans}^{0:t-1} \lambda^{0:t-1} \tau^{0:t-1} \beta^{0:t} \pi) \\
\propto \left| \begin{array}{l}
P(B^t|zx_{dir}^{0:t} zx_{dist}^{0:t} b^{0:t-1} m_{rot}^{0:t-1} m_{trans}^{0:t-1} \lambda_{pred}^{0:t-1} \tau_{pred}^{0:t-1} \beta_{pred}^{0:t} \pi_{pred}) \\
\times P(B^t|zy_{dir}^{0:t} zy_{dist}^{0:t} b^{0:t-1} m_{rot}^{0:t-1} m_{trans}^{0:t-1} \lambda_{prey}^{0:t-1} \tau_{prey}^{0:t-1} \beta_{prey}^{0:t} \pi_{prey}) \\
\times P(B^t|zh_{dir}^{0:t} zh_{dist}^{0:t} b^{0:t-1} m_{rot}^{0:t-1} m_{trans}^{0:t-1} \lambda_{home}^{0:t-1} \tau_{home}^{0:t-1} \beta_{home}^{0:t} \pi_{home}),
\end{array} \right.
\end{array}
\tag{6.15}
$$

where $Z_{lf}^{0:t}$ is the set of all observation variables as defined in equation 5.15. From the above global distribution for the behaviour variable, a value for the present-time behaviour variable is drawn.

The estimation question is then asked to each elementary filter, with the addition of the drawn value b^t in the right part of the question. For the predator filter, for example, we have:

$$
P(X_{dir}^t X_{dist}^t | zx_{dir}^{0:t} zx_{dist}^{0:t} b^{0:t} m_{rot}^{0:t-1} m_{trans}^{0:t-1} \lambda_{pred}^{0:t-1} \tau_{pred}^{0:t-1} \beta_{pred}^{0:t} \pi_{pred}).
$$

After estimation, the motor question is asked to each elementary filter. For the predator filter, the motor question is:

$$
P(M_{rot}^t M_{trans}^t | zx_{dir}^{0:t} zx_{dist}^{0:t} b^{0:t} m_{rot}^{0:t-1} m_{trans}^{0:t-1} \lambda_{pred}^{0:t} \tau_{pred}^{0:t} \beta_{pred}^{0:t} \pi_{pred}).
$$

The combination of answers from all elementary filters is then:

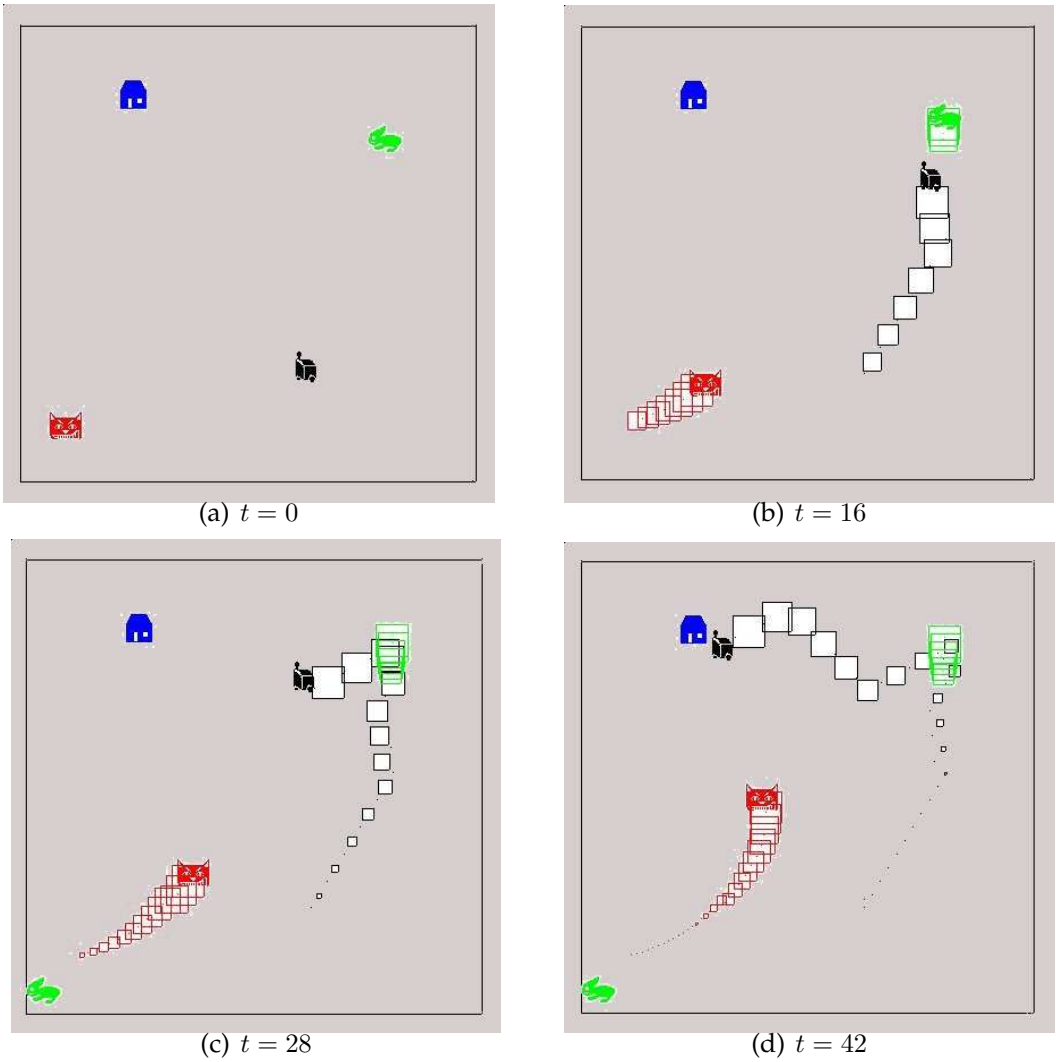
$$
\begin{array}{l}
P(M_{rot}^t M_{trans}^t | z_{lf}^{0:t} b^{0:t} m_{rot}^{0:t-1} m_{trans}^{0:t-1} \lambda^{0:t} \tau^{0:t} \beta^{0:t} \pi) \\
\propto \left| \begin{array}{l}
P(M_{rot}^t M_{trans}^t | zh_{dir}^{0:t} zh_{dist}^{0:t} b^{0:t} m_{rot}^{0:t-1} m_{trans}^{0:t-1} \lambda_{home}^{0:t} \tau_{home}^{0:t} \beta_{home}^{0:t} \pi_{home}) \\
\times P(M_{rot}^t M_{trans}^t | zx_{dir}^{0:t} zx_{dist}^{0:t} b^{0:t} m_{rot}^{0:t-1} m_{trans}^{0:t-1} \lambda_{pred}^{0:t} \tau_{pred}^{0:t} \beta_{pred}^{0:t} \pi_{pred}) \\
\times P(M_{rot}^t M_{trans}^t | zy_{dir}^{0:t} zy_{dist}^{0:t} b^{0:t} m_{rot}^{0:t-1} m_{trans}^{0:t-1} \lambda_{prey}^{0:t} \tau_{prey}^{0:t} \beta_{prey}^{0:t} \pi_{prey}).
\end{array} \right.
\end{array}
\tag{6.16}
$$

Simulation Results

Figure 6.2 shows a sequence of snapshots for a simulation run using the Bayes filter described above. One behaviour is selected at each time instant, and Figure 6.3 displays the drawn behaviour values for all time instants of the simulation run described in the figure.

Initially, the robot selects behaviour Chase (Figure 6.2(b)), then changes transiently to Escape and Motionless behaviour (Figure 6.2(c)). Finally it selects Homing (Figure 6.2(d)).

The situation at time step $t = 28$, pictured in snapshot 6.2(c), is very interesting as it illustrates different motor patterns (escaping, going towards and motionless) and how the robot switches from one pattern to the other. Figure 6.4 shows behaviour and rotation speed fusions at three different time instants: $t = 27$, $t = 28$ and $t = 30$. On the left, the curves display the probability distributions for the behaviour variable in each elementary filter and the result of fusion. On the right side, rotation speed proposals from each elementary filter and the resulting fusion are shown.



(a) $t = 0$ (b) $t = 16$

(c) $t = 28$ (d) $t = 42$

Figure 6.2: Sequence of snapshots for simulation of the illustrative example.

It can be seen that at time instant $t = 27$, in the left part of Figure 6.4(a), the predator filter discourages the choice of chase behaviour, while all other filters recommend it. Actually, the robot was chasing, and as no other stimuli came from the environment, other filters indicate continuing the previous behaviour. The predator filter, however, can detect a predator near enough to recommend any self-defence-related behaviour and disapprove chase behaviour. Even if fusion indicates a clear preference for continuing the chase, the drawn value for the behaviour variable is escape. As a consequence, the rotation speed curves (in the right part of Figure 6.4(a)) are uniform for all filters, except the predator filter, which indicates the preferred direction to follow to escape the predator.

Because of the approaching predator, in the next time step ($t = 28$), the predator filter enhances its recommendation for a self-defence behaviour and at the same time

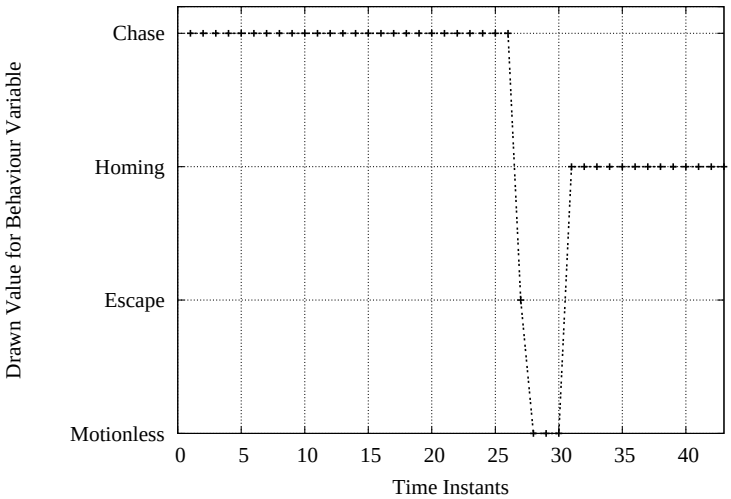


Figure 6.3: Drawn values for behaviour variables for simulation of the illustrative example.

rejects chasing, as can be seen in the left part of Figure 6.4(b). The resulting fusion now favours self-defence behaviours, especially homing. The drawn value selects Motionless behaviour, and the rotation speed curves shown in the right part of the same figure indicate that only the predator filter proposes a curve different from uniform, which recommends null rotation.

The next change in behaviour happens at $t = 30$, the situation shown in Figure 6.4(c). In the left of the figure, we can see that the filter proposals for behaviour selection and the resulting fusion are not very different from those at time step $t = 28$, but the drawn value for behaviour variable now is homing. In the right part of the figure, it is shown that the prey filter has no suggestion for rotation speed, while the predator filter recommends against the predator direction. On the other hand, the home filter highly recommends the home direction. The fusion result indicates the home direction as more probable.

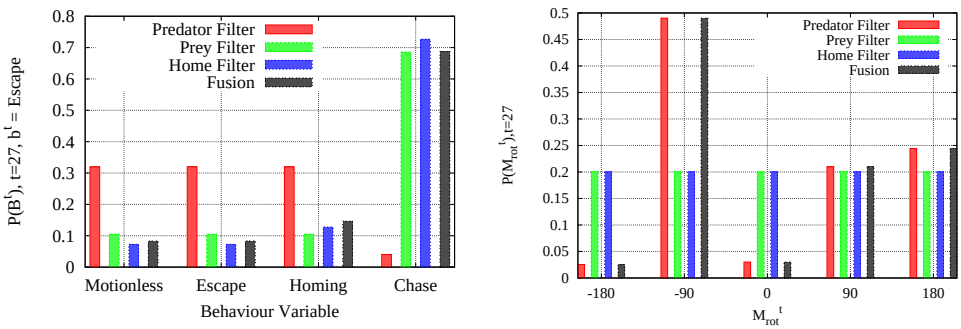
From now on, homing behaviour becomes stronger as the distance from the robot to home decreases. This behaviour is selected steadily until the robot arrives home, and the simulation is stopped, as shown in Figure 6.3.

6.1.4 Additional Main Assumptions

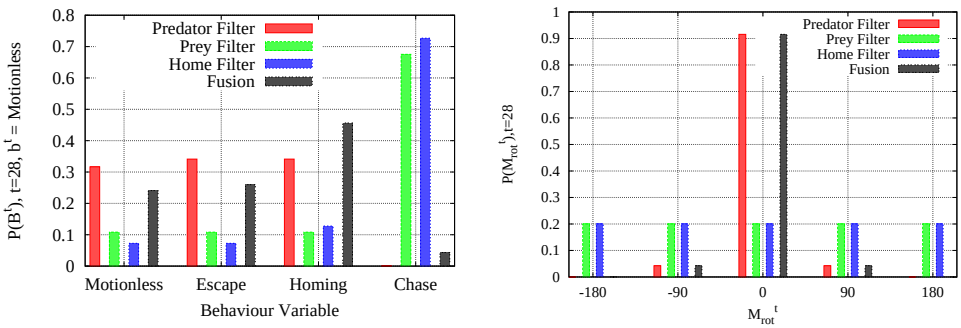
- Motor commands can be assembled in well-identified patterns, called behaviours.

  Robot displacement and actions in the environment result from coordinated control of actuators, and it is plausible to group motor commands into patterns, called here behaviours.

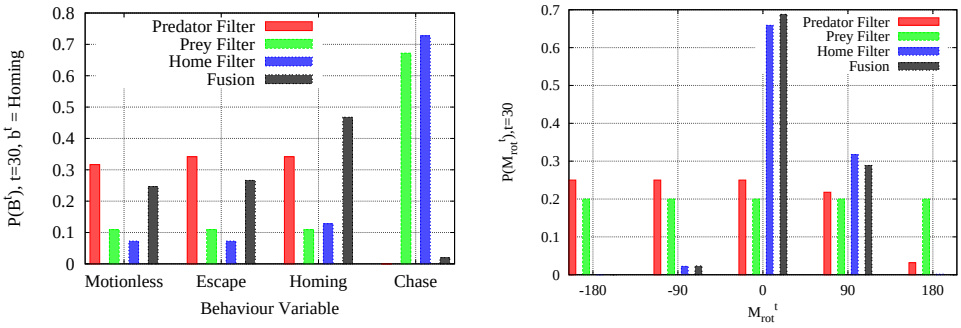
- The strategy for behaviour selection is individual for each elementary filter. Filter



(a) $t = 27$. Drawn value for behaviour variable is Escape.



(b) Drawn value for behaviour variable is Motionless. $t = 28$



(c) Drawn value for behaviour variable is Homing. $t = 30$

Figure 6.4: Behaviour selection (at left) and rotation speed (at right) at three different time instants. Behaviour selection curves show the answers of all elementary filters to the behaviour question and the resulting fusion. Rotation speed curves show the answers of all elementary filters to the motor question and the resulting fusion: the behaviour value employed in the motor question is the drawn value at each time instant.

proposals are combined for a global selection.

Selection of the most relevant motor pattern for the present time step derives from the combination of elementary filter proposals. Each elementary filter rates the adequacy of each possible behaviour at the present time step in terms of its estim-

ation of its own state variables.

- The choice of present behaviour is influenced by robot and environment features, and by the robot's specified behaviour.

  Inside each elementary filter, the behaviour selection model defines how the filter state variables are related to behaviour selection, considering the specified behaviour and state variable semantics. This ensures reactivity in the robot's resulting behaviour.

- Behaviour selection depends on behaviour at the previous time step.

  The behaviour selection model also shows a dependence on time: to ensure persistence and continuity in time when executing a specific motor pattern, behaviour selection at the present time depends on the behaviour selected in the previous time step.

- Only one behaviour is active at a time.

  Robot actuators can execute only one motor pattern at any time instant, and this implies that behaviour variables must have known values when the motor command question is asked of the elementary filters.

6.2 Requirements Analysis

We have seen in the previous chapter that employing elementary filters drastically reduces complexity in the Bayes filter. We would like to verify, in the following sections, the effect of adding the behaviour model on space and time complexities.

6.2.1 Space Complexity

Only elementary filter joint distributions must be kept in memory, and the number of cells for storing each term is as follows. N_b denotes the number of behaviour variables, which are considered to have the same number of cases n as all other variables.

- A priori from previous time step: $P(S_i^{j-1}|\pi_i)$: $n^{N_{s_i}}$ cells;
- Dynamic model: $P(S_i^j|S_i^{j-1}M^{j-1}\pi_i)$: $n^{2N_{s_i}+N_m}$ cells;
- Sensor model: $P(Z_i^j|S_i^j\pi_i)$: $n^{N_{z_i}+N_{s_i}}$ cells;
- Behaviour model: $P(\beta_i^j|B^jS_i^jB^{j-1}\pi_i)$: $2n^{2N_b+N_{s_i}}$ cells;
- Motor command model: $P(\lambda_i^j|M^jB^jS_i^jM^{j-1}\pi_i)$: $2n^{2N_m+N_{s_i}+N_b}$ cells.

Answers to questions must also be stored. These are the prediction, behaviour, estimation and motor questions for each elementary filter, and the global questions of the behaviour and motor commands. Adding all these values we have:

$$\mathcal{M} = \left| \begin{array}{l} \sum_{i=1}^{N_i} \left[\begin{array}{l} n^{2N_{s_i}+N_m} + n^{N_{z_i}+N_{s_i}} \\ +2n^{2N_b+N_{s_i}} + 2n^{2N_m+N_b+N_{s_i}} \\ +3n^{N_{s_i}} \end{array} \right] \\ +(N_i+1)(n^{N_m} + n^{N_b}). \end{array} \right. \tag{6.17}$$

Compared with the previous version, the additional terms in the above equation are because of the behaviour model, storage of the behaviour question and the new dependence in the motor model, where the behaviour variable is included.

For our illustrative example, 161 393 probability values must be kept in memory. In the previous chapter, there were 111 844 probability values; this means that addition of the behaviour selection mechanism was responsible for an increase of almost 50 000 probability values to be stored. Actually, if conditional independence in both behaviour and motor command models is considered, the required memory space is smaller.

The amount of space necessary to keep this Bayes filter in memory has the upper bound:

$$\mathcal{M} = O\left(\begin{array}{l} n^{2max(N_{s_i})+N_m} + n^{max(N_{z_i}+N_{s_i})} + n^{2N_b+max(N_{s_i})} \\ +n^{2N_m+N_b+max(N_{s_i})} + N_i \times (n^{N_m} + n^{N_b}) \end{array} \right). \tag{6.18}$$

Comparing this expression with equation 5.18:

$$\mathcal{M} = O\left(\begin{array}{l} n^{2max(N_{s_i})+N_m} + n^{max(N_{z_i}+N_{s_i})} \\ +n^{max(N_{s_i})+2N_m} + N_i \times n^{N_m} \end{array} \right), \tag{6.19}$$

we can see that addition of the behaviour variables and model does not change the green term, related to the dynamic model, nor the blue term for the observation model. However, it does influence the motor model term (in brown): this is because of the dependence on behaviour variables in the motor model. In addition, the dependence on the number of elementary filters, given by the black term, is now augmented by the number of behaviour variables, as behaviour questions for each elementary filter must be stored.

A new term appears in equation 6.18, which concerns the behaviour model, in red. We can see that the amount of memory scales with the biggest subset of state and observation variables and with the number of elementary filters.

Applying the sensor fusion schema wherever possible, as presented in Chapter 3, it is possible to reduce the influence of the observation variables, as shown in the space complexity analysis of the previous chapter. The amount of required memory when employing sensor fusion is:

$$\mathcal{M} = \left| \begin{array}{l} \sum_{i=1}^{N_i} \left[\begin{array}{l} n^{2N_{s_i}+N_m} + N_{z_i} n^{N_{s_i}} \\ +2n^{2N_b+N_{s_i}} + 2n^{2N_m+N_b+N_{s_i}} \\ +3n^{N_{s_i}} \end{array} \right] \\ +(N_i+1)(n^{N_m}+n^{N_b}), \end{array} \right. \tag{6.20}$$

and the upper bound becomes:

$$\mathcal{M} = O\left(\begin{array}{l} n^{2max(N_{s_i})+N_m} + max(N_{z_i}) n^{max(N_{s_i})} + n^{2N_b+max(N_{s_i})} \\ +n^{2N_m+N_b+max(N_{s_i})} + N_i \times (n^{N_m}+n^{N_b}) \end{array} \right). \tag{6.21}$$

Comparing with the previous chapter's upper bound, shown in equation 5.21:

$$\mathcal{M} = O\left(\begin{array}{l} n^{2max(N_{s_i})+N_m} + max(N_{z_i}) n^{max(N_{s_i})} \\ +n^{2N_m+max(N_{s_i})} + N_i \times n^{N_m} \end{array} \right). \tag{6.22}$$

we can see that the dependence on the number of observation variables is linear, and there is no change in the influence of behaviour variables or the behaviour model compared with the above analysis.

It is possible to simplify the space complexity expression even more if some assumptions about relations between numbers of variables are considered.

As stated in the previous chapter, it is reasonable to assume that:

$$N_i \ll n^{max(N_{s_i})}, \tag{6.23}$$

and then the space complexity no longer depends on the number of elementary filters.

Considering the role of behaviour variables as selector switches of motor patterns, it is acceptable to assume that the number of cases for all behaviour variables is smaller than the number of cases for all motor variables, or:

$$n^{N_b} \lll n^{N_m}. \tag{6.24}$$

In this case, the brown term in equation 6.18 becomes:

$$n^{2N_m+N_b+max(N_{s_i})} = n^{2N_m+max(N_{s_i})}. \tag{6.25}$$

For the same reasons, it is possible to consider that the number of cases of all behaviour variables is smaller than the number of cases of the biggest subset of state variables ($n^{N_b} \ll n^{max(N_{s_i})}$). The red term in equation 6.18 is then no longer relevant to space complexity.

The above assumptions lead to the conclusion that the space complexity does not change with the addition of behaviour variables for most usual applications in robotics.

6.2.2 Time Complexity

The sequence of questions asked to the filter, described in section 6.1.2, determines the computational effort for one time step.

- Prediction in each elementary filter, as in equation 6.4, requires $4n^{2N_{s_i}} - n^{N_{s_i}}$ operations.

$$
\begin{aligned}
&P(S_i^t | z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i) \\
&\quad \propto \sum_{S_i^{t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(m^{t-1} | \pi_i) \times P(\lambda_i^{t-1} | m^{t-1} S_i^{t-1} b^{t-1} m^{t-2} \pi_i) \\ \times P(S^{t-1} | z_i^{0:t-1} b^{0:t-1} m^{0:t-2} \lambda_i^{0:t-2} \beta_i^{0:t-1} \pi_i) \end{array} \right] .
\end{aligned}
$$

- Behaviour Selection in each elementary filter, as shown in equation 6.5, requires $4n^{N_{s_i}+N_b} - n^{N_b}$ operations.

$$
\begin{aligned}
&P(B^t | z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \pi_i) \\
&\quad \propto \sum_{S_i^t} \left[\begin{array}{l} P(z_i^t | S_i^t \pi_i) \\ \times P(B^t | \pi_i) \times P(\beta_i^t | B^t b^{t-1} S_i^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i) \end{array} \right] .
\end{aligned}
$$

- Computing Global Behaviour Selection, as in equation 6.6, requires $N_i n^{N_b}$ operations.

$$
\begin{aligned}
&P(B^t | z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t} \pi_{bs}) \\
&\quad \propto \prod_{i=1}^{N_i} P(B^t | z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \pi_i).
\end{aligned}
$$

- Estimation in each elementary filter, according to equation 6.7, requires $3n^{N_{s_i}}$ operations.

$$
\begin{aligned}
&P(S_i^t | z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \pi_i) \\
&\quad \propto \left| \begin{array}{l} P(z_i^t | S_i^t \pi_i) \\ \times P(b^t | \pi_i) \times P(\beta_i^t | b^t b^{t-1} S_i^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i). \end{array} \right.
\end{aligned}
$$

- Motor Commands in each elementary filter, according to equation 6.8, require $3n^{N_{s_i}+N_m} - n^{N_m}$ operations.

$$\begin{aligned} & P(M^t|z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t}\beta_i^{0:t}\pi_i) \\ & \propto \sum_{S_i^t} \left[\begin{array}{l} P(M^t|\pi_i) \times P(\lambda_i^t|M^t S_i^t b^t m^{t-1}\pi_i) \\ \times P(S_i^t|z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\pi_i) \end{array} \right] . \end{aligned}$$

- Global Motor Commands, as expressed in equation 6.9, involve $N_i n^{N_m}$ operations.

$$\begin{aligned} & P(M^t|z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\pi_{bs}) \\ & \propto \prod_{i=1}^{N_i} P(M^t|z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t}\beta_i^{0:t}\pi_i). \end{aligned}$$

Considering all the above calculations, the expression for the number of required operations becomes:

$$\mathcal{C} = \sum_{i=1}^{N_i} \left[\begin{array}{l} 4n^{2N_{s_i}} + 2n^{N_{s_i}} \\ +4n^{N_b+N_{s_i}} + 3n^{N_m+N_{s_i}} \end{array} \right]. \tag{6.26}$$

Computation time has clearly increased compared with the previous version, as the behaviour question asked of each filter requires a summation over the state space (as can be seen in equation 6.5). This is the computational cost of adding the new functionality of a behaviour selection mechanism.

For our example, the number of necessary operations at each time step is 17 078, compared with 11 256 in the previous chapter.

The time complexity has the upper bound:

$$\mathcal{C} = O\left(n^{2max(N_{s_i})} + n^{N_m+max(N_{s_i})} + n^{N_b+max(N_{s_i})} \right). \tag{6.27}$$

Comparing with equation 5.25 from the previous chapter:

$$\mathcal{C} = O(n^{2max(N_{s_i})} + n^{max(N_{s_i})+N_m}). \tag{6.28}$$

we see that the green part, related to uncertainty propagation from time step $t-1$ to time step t, and the brown term, the consequence of asking the motor question, are unchanged. The new red term in equation 6.27 results from asking the behaviour question.

The dependence on the number of observation variables is added when sensor fusion is applied, and the number of required operations is:

$$\mathcal{C} = \sum_{i=1}^{N_i} \left[\begin{array}{l} 4n^{2N_{s_i}} + (2N_{z_i} - 1)n^{N_{s_i}} \\ +4N_{z_i} n^{N_b+N_{s_i}} + 3n^{N_m+N_{s_i}} \end{array} \right]. \tag{6.29}$$

Time complexity now has the upper bound:

$$\mathcal{C} = O\left(\begin{array}{l} n^{2max(N_{s_i})} + max(N_{z_i})n^{max(N_s)} \\ +n^{N_m+max(N_{s_i})} + max(N_{z_i})n^{N_b+max(N_{s_i})} \end{array} \right). \tag{6.30}$$

Comparing with the complexity of employing sensor fusion in the previous chapter, in equation 5.28:

$$\mathcal{C} = O(n^{2max(N_{s_i})} + max(N_{z_i})n^{max(N_s)} + n^{max(N_{s_i})+N_m}), \tag{6.31}$$

we can see that the new term (in red) related to the behaviour questions depends on the biggest subset of observation variables.

Considering that the number of cases of all behaviour variables is smaller than the number of cases of the biggest subset of state variables ($n^{N_b} \ll n^{max(N_{s_i})}$), the red term becomes equal to the blue one, and the complexity expression becomes similar to the one in the previous chapter, as stated in equation 5.28.

6.2.3 Programming

To decide which motor pattern to apply, all environment features must be taken into account, and for that reason, the behaviour selection variable is common to all elementary filters.

The global joint distribution shown in equation C.7 establishes the relation between elementary filter state variables and the behaviour variable: all assumptions in relation to variable dependencies are present in the joint distribution, but for some applications, further simplifications are possible. To benefit from the advantages of elementary filters, the behaviour model in each elementary filter includes dependence only on the filter state variables, and utilisation of the coherence-based fusion form when describing the behaviour selection model allows posterior fusion of elementary filter proposals.

The behaviour selection model then implements the selection strategy in each elementary filter, based on its own state variables, and consequently knowledge about state variables in other filters is not available. Taking into account the posterior fusion of elementary filter proposals, it is more efficient to apply proscriptive programming to build the behaviour selection model.

Proscriptive programming emphasises the need to prohibit forbidden values more than to recommend desired values. It is highly desired in a sensor fusion schema, and as coherence-based fusion gives a similar result, it is also useful for both motor commands and behaviour selection models. The benefits of programming proscriptively using probability distributions, especially when they are combined by a fusion method, was discussed by Lebeltel [1999, page 178].

Levels are chosen for classifying behaviours for a given situation: highly recommend, recommended, not relevant, unwise, strictly prohibited. Each situation is defined by associating a probability value for each possible behaviour selected in the previous time step and the possible values of the state variables at the present time. The strictly

prohibited probability value must be as small as necessary to prevent its selection, even if another filter recommends it.

Instead of having a very big number of probability values to adjust (approximately equal to the amount of memory necessary to store the behaviour model in memory, shown above), only five parameters must be tuned: the probability values used in Table 6.1.

Learning

The learning process can be simplified in this proposed system. Actually, as each elementary filter behaviour and motor model is programmed separately, it is possible to design separate learning processes for each filter model.

Diard and Lebeltel [2000] proposed a method for learning behaviour combinations and behaviour transition tables. A khepera robot is guided using a joystick. While it is learning the behaviour combination distribution, situation recognition is applied to identify parameters in a Laplace distribution. However, while learning the behaviour transition table, the robot is guided using not motor commands but behaviour values at each situation.

Video game character behaviour is tuned by real-time learning [Le Hy et al., 2004]. Behaviour transition tables of probabilities (given a priori and learnt) are compared with native bots, and results show that both perform better that native bots, with learnt aggressive bots showing very superior performance.

Both methods can be applied to learn behaviour transition tables using the method proposed here, with the advantage of separate learning models for each filter. Each filter actually represents a different context of learning.

How can we ensure that fusion of behaviour proposals based on models learnt separately is coherent with the specified behaviour? This is an important question remaining to be verified.

6.3 Discussion

6.3.1 Robotics

An autonomous robot is assumed to be able to decide what to do to fulfil its goals. Deciding what to do next means choosing, among possible actions, one that is compatible with the present configuration of the environment. The problem of making a decision is called the *action selection*, or the *behaviour coordination* problem.

We discuss here two main points regarding action selection mechanisms: the granularity of action definition, and the usual requirements for an action selection mechanism.

Definition of Action and the Action Selection Problem

From a robotics point of view, the action selection problem definition is related to optimisation: the choice of an action consists of maximising its utility towards the robot's specified goals or global behaviour while simultaneously minimising its costs. Different approaches in robotics lead to different interpretations of what are *goals*, *utility* and *costs*, so the definition above is nuanced according to the approach.

Reviews of the action selection problem in robotics have been presented by Maes [1990], and Pirjanian [1999]. On the other hand, the reviews by Tyrrell [1993], and Prescott et al. [1999] cover other related domains such as ethology and neuroscience. Reinforcement learning applied to selection between conflicting and heterogeneous goals was treated by Humphrys [1997]. There, a distinction was made between selecting an action to accomplish a unique goal and choosing between conflicting goals.

This difference emphasises a basic difficulty in defining the action selection problem: the granularity of actions. An action can vary from a particular motor command (for example, *turn right*) to more complex combination of motor controls (for example, *return home*).

In a practical approach, Bryson [2001] discussed the action selection problem as **when** to do **what** and **how**. The complexity of actions (**what**) represents a trade-off in the complexity of **how** (coordination of motor commands to execute actions) and the complexity of **when** (the action selection problem). If the **what** (actually, the actions) consists of very basic motor commands, the definition of **how** to execute an action is rather simple, but the **when** becomes more complicated. This is the situation in the system proposed in chapter 5, where the decision is made directly over motor command variables.

In the mechanism proposed in this chapter, we aim to have more complex actions than motor commands. In other words, we want a more complex **what** for a simpler and more effective **when**, even if **how** is slightly more complicated.

In the illustrative example, the granularity of the behaviour variables is determined by the motor patterns and it can happen that similar motor patterns are applied to different goals: *go towards the prey* is a motor pattern very near to *go towards home*. In this case, behaviour variables are more related to goals (which may conflict) than to actions towards a unique goal. In contrast, *escape* and *motionless* are different actions for accomplishing the same goal (not to be captured by the predator).

Baherend's model for animal action selection, described by Tyrrell [1993] and Prescott et al. [1999], includes relations between primary goals (called system-level goals) and low-level motor actions (called actions). This model was not conceived to be used in robotics, but an extension called the behaviour system approach (see Timberlake and Lucas [1989], Timberlake and Silva [1995] and Timberlake [2001] for details) was applied in a commercial robot by Arkin et al. [2001]. Employing this model as basis for a hierarchy, it is possible to discern goals, strategies and motor patterns.

The inclusion of a behaviour hierarchy in our approach can be achieved in different ways and is the subject of ongoing work.

Action Selection Mechanism Requirements

Some of the basic requirements for an action selection mechanism proposed in the literature [Maes, 1990; Tyrrell, 1993; Pirjanian, 1999] are presented and discussed in this section. It is important to remark that some of them are imprecisely defined, or represent subjective concepts, while others are conflicting or incompatible. For these reasons, these requirements should be considered more as guidance for designing an action selection mechanism than as hard evaluation criteria to be fully met.

Goals

This is one of the main requirements of an action selection mechanism: to take into account the goals of the specified robot behaviour and to select actions that contribute to the accomplishment of these goals.

It is important to differentiate a behaviour goal and a goal position. In robotics, navigation usually includes an environment map, with associated problems of localisation, planning and plan execution, with the necessary definition of a goal position: the desired robot position at the end of path planning. More generally, a specified environment state must be achieved. Here, the behaviour goal concerns internal motivations: chasing is related to feeding needs, as escaping involves self defence, for example.

Often, complex specified behaviours present several simultaneous goals: in our example, the robot must escape from predators and chase prey. A very interesting discussion about types of goal in the context of planning can be found in Dean et al. [1995].

In our approach, the behaviour specification defines all task goals, and the values of the behaviour variables are defined from this specification. Behaviour specification also helps to identify relations and dependencies between state variables and behaviour variable values, as well as associated motor patterns.

When the behaviour model is being programmed, goals are built-in in the dependence on the state variables. In our illustrative example, one of the robot goals is to avoid being caught by the predator. This is made explicit by the behaviour value '*Escape from predator*', which is tightly related to the predator distance in the behaviour model, and it implies a motor pattern of not going in the predator direction.

In the behaviour model, behaviours that allow pursuit of a relevant goal receive a probability value of recommendation in each elementary filter. Simultaneous goals are combined during fusion of the filter propositions, and if one behaviour favours more than one goal, it will have a higher probability value.

Tyrrell [1993, p. 90] discussed goals as sub-problems to be solved by the action selection mechanism and among the different types of sub-problems, with the definition of prescriptive and proscriptive ones. *Prescriptive* describes the sub-problems requiring that some actions be carried out, while *proscriptive* describes sub-problems requiring that some actions must not be carried out.

These definitions are very relevant to behaviour model programming: as behaviours are associated with one or several goals of the robot's specified behaviour, a clear distinction between prescriptive and proscriptive behaviour is very helpful.

Combinations of elementary filter proposals are very near the compromise actions discussed in Tyrrell [1993, p. 170] and Crabbe [2004, 2005]. Tyrrell remarked that compromise actions (solving more than one goal) should be preferable to an action benefiting only one goal. This was established as one of the requirements to be fulfilled by our action selection mechanisms.

Crabbe [2005] investigated the nature of compromise actions under different simulated environmental conditions, with both prescriptive and proscriptive goals.

When analysing simulated results, Crabbe defined two levels where compromise models are different: compromise actions, which are low-level motor command actions; and compromise behaviours at higher levels, which may have different effects on lower level actions. Blending the compromise model, where compromise actions are the result of blending actions best for each goal, was considered to take place at the low level of motor command actions. In voting on compromises, a collection of experts vote on actions favouring goals, and a central module selects the action with the most votes. This happens at a higher level, where actions are more abstract.

In our system, these two levels are clearly separated: motor commands are on the lower level, and behaviour (motor patterns) are at a higher level. Compromise models are also present: voting in global behaviour fusion (as can be seen in the illustrative example in Section 6.1.3) and blending in motor command fusion. Both models are actually implemented using the same technique, and the difference results from the chosen parametric forms: probability tables in the case of behaviour fusion allows voting-like fusion; while normal functions in the case of motor commands indicate blending fusion.

Crabbe also concluded that compromise actions seem to be more useful at higher levels of action abstraction, as at low levels, they can bring limited benefits when proscriptive goals are used. We have successfully implemented obstacle avoidance employing proscriptive programming [Koike et al., 2003a,b], where the same compromise action model described in this chapter was employed.

Although our work does not include any kind of optimisation in action selection, in contrast to Crabbe [2004, 2005], common findings can lead to stronger evidence on utilisation of proscriptive goals and compromise in action selection.

Persistence

When several goals are present, a robot can alternate between two or more choices repetitively. These oscillations are undesirable because they are ineffective, and neither goal is ever reached.

A specific action is usually selected based on the intensity of perception cues. If the goal associated with this action is not yet satisfied, it is not desired to change to another goal if the intensity of perception stimuli decreases.

In these situations, persistence is desired as a way to ensure that an action aiming to fulfil one of the goals will be given a certain preference until the goal is reached.

Persistence is included in our framework by means of the dynamics in the behaviour model. Actually, B^t depends on B^{t-1}, and so a bigger probability is attributed in the behaviour model when $B^t = B^{t-1}$.

On the other hand, it is important not to keep persistence exclusive, otherwise behaviour transitions will rarely happen. With proscriptive programming for the behaviour selection model, this means having intermediary values of probability for some behaviours, such as not related or unwise, as described in Section 6.1.3. Fusion of filter proposals will allow these intermediary behaviours to be proposed and selected by other filters, causing a transition.

Reactivity

When trying to orient selected actions towards a goal, it is still necessary to react quickly to hazardous situations and unexpected events. The current goal is then discarded to react to this new situation. This requirement is called reactivity.

Most of the time, the reactivity requirement is related to fast decisions, as stated by Pirjanian [1999].

In our filter, reactivity is ensured by the dependence on state variables in the behaviour selection model. As observation values are also taken into account when asking behaviour question, state variables are updated when employed for calculating behaviour propositions in each elementary filter.

Opportunism

Whenever the environment situation changes and a selected action towards a goal is no longer interesting, the action selection mechanism should be opportunist and chose another goal or another action. Opportunism is a requirement frequently cited when modelling animal action selection [Tyrrell, 1993; Tu, 1996].

This concept is close to the reactivity requirement described above; the main difference lies in the compulsory nature of reactivity versus convenience in the case of opportunism. This implies also that reactions are fast, impulsive answers to stimuli, while opportunistic decisions result from weighing different situations and choosing the most relevant.

In our behaviour model, an opportunistic situation generates a transition from a persistent behaviour. Persistence is programmed in the behaviour model as the default suggestion of all filters: this proposal is based only on the B^{t-1} values. Opportunistic behaviour depends on state variables, and is usually proposed by only one filter: the recommendation strength and eventual proscription of other filters will then determine if the opportunistic behaviour is dominant after fusion of filter suggestions.

Planning

In some complex behaviours, a sequence of actions must be executed to fulfil a goal: planning means identifying this sequence beforehand and selecting actions in the right order and at appropriate instants.

In our approach, it is the dependence between B^t and B^{t-1} that allows explicit sequences in behaviour choices. A transition from one behaviour to another can thus be made explicit with additional dependence on environment triggers (by means of state variables).

As discussed in Chapter 4, the Bayes filter proposed here does not consider future time to decide the next action. Although prespecified sequences of actions can be programmed in the behaviour selection model, there is no planning in the sense of the policy calculation in the POMDP approach, as discussed in Section 4.3.1. Behaviours are selected based on the past and present situations defined by state variables, but future consequences, rewards and costs of selecting a behaviour are not taken into account.

It is clear that it is not possible to achieve fully all these requirements: persistence and opportunism, for example, may conflict in some situations. Only by compromise is it possible to create a system that is both reactive and goal directed.

Evaluation of how these requirements are attained is not precise, and there is no widely accepted methodology. Tyrrell [1993] proposed a simulated environment, and within the scope of this environment, a procedure for evaluating and comparing several action selection mechanisms was introduced. Most robotics studies provide real system results as evidence that their proposed action selection mechanism is operational. However, as results are obtained employing different physical robots, assorted computational platforms and uncontrolled experimental conditions, it is not possible to compare results quantitatively.

Behaviour variables: States or Observations?

As cited in Chapter 4.3.1, motor variables are considered a special kind of state variables with known past values. Their influence over system time complexity was also discussed.

Similarly, when the behaviour selection question is asked in each elementary filter, and also when answers are later combined in a global distribution over behaviour variables, these variables are being estimated. Behaviour variables become observations when their values are drawn from this global distribution.

It is possible to verify, in the time complexity expression shown in equation 6.27, that exponential dependence terms exist for both motor (in brown) and behaviour (in red) variables.

In addition to its influence over time complexity, the duality of behaviour variables has further consequences. When the estimation question is calculated, as shown in equation 6.7, behaviour variables have associated known values: they are employed in a sensor fusion schema in the expression answering the state estimation question, together with observations. The behaviour model then affects the accumulated knowledge expressed by the prediction, and it consequently influences state estimation.

A major consequence of this additional influence is that state variable values will be strengthened or attenuated according to their relation to the drawn values of behaviour variables.

Furthermore, behaviour variable values are drawn from global distributions: this implies that state variables in an elementary filter can be highly influenced by behaviour

values from another elementary filter.

To exemplify how the choice of behaviour influences state variables in each elementary filter, let us examine the transition between motionless and homing behaviour in our example. Figure 6.5 shows prediction and estimation of home filter state variables at two time instants: $t = 30$ and $t = 31$.

The selected behaviour at time instant $t = 30$ is motionless, and this behaviour is not directly related to the state variables in the home filter. The prediction of home distance and direction shown in Figure 6.5(a) indicates that home is not close and in a direction around zero in relation to the robot heading. Estimation in Figure 6.5(b) updates this information using observations ($ZH_{dir} = 90$ and $ZH_{dist} = 3$) and behaviour model influence (with selected behaviour b^t equal to motionless). Home is then better localised, mostly influenced by observations, because the home behaviour model is not directly concerned with motionless behaviour.

For time instant $t = 31$, however, there is a much bigger difference between prediction and estimation. The main reason is that the selected behaviour is homing, a behaviour directly concerned with state variables in the home filter. In the estimation phase, observations are unchanged ($ZH_{dir} = 90$ and $ZH_{dist} = 3$), but the behaviour model (which now assumes that if homing was selected, then the home distance is small) also updates the prediction. As a result, the estimation of the home distance is biased by the selected behaviour and indicates that home is near the robot, as can be seen in Figure 6.5(d).

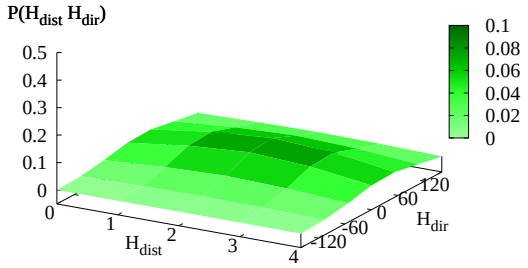
It is important to remark that the same type of influence happens with the selected values of motor commands, but in the prediction expression. It is clear in equation 6.4 that motor command values from the previous time step (m^{t-1}) influence estimation in the previous time step, and prediction of state variable values will be strengthened or attenuated according to their relation to the drawn values of motor command variables.

This major consequence of deciding a variable value may appear to be a degradation of state estimation quality, as a *false* observation is added in fusion with observations (for behaviour variables) or with previous time estimation (for motor command variables).

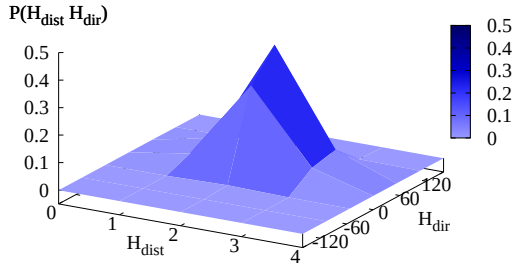
Our argument is that whenever a decision is taken over a variable, whether a behaviour or a motor command variable, state variables may become inconsistent with present assumptions about the environment. The influence of the drawn behaviour value over the state variables is a way to propagate the selected value assumptions over other variables in the filter.

On the other hand, selected values can be ignored and not taken into account: uncertainty regarding these variables is then propagated, as described in Section 4.3.1. The distribution over these variables will then be present in both prediction and estimation steps in a probabilistic way.

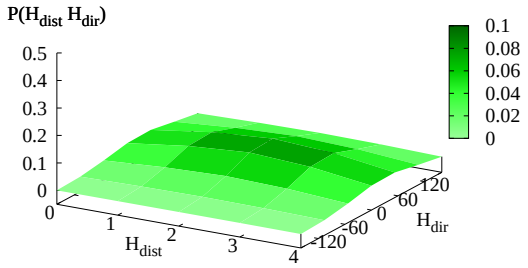
Following the mathematical rigour of the joint distribution definition and utilisation, the influence of these variables over the state variables is inevitable, either as selected values or as probability distributions. This influence is an inevitable consequence of behaviour selection being dependent on the state variables.



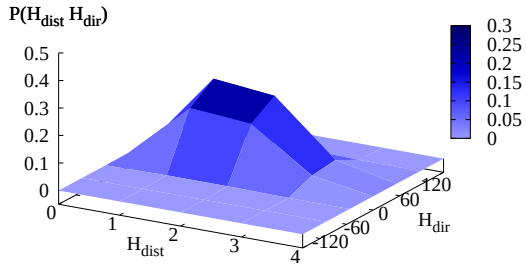
(a) Prediction of Home States at time instant $t = 30$



(b) Estimation of Home States at time instant $t = 30$



(c) Prediction of Home States at time instant $t = 31$



(d) Estimation of Home States at time instant $t = 31$

Figure 6.5: States in the home filter for time instants $t = 30$ and $t = 31$.

6.3.2 Cognition

The cognitive interpretations of the mathematical hypotheses added in this chapter (see decomposition in equation 6.1) are the following:

- The Central Nervous System uses a repertoire of independent sensory-motor modules (synergies). A new type of internal state variables (variable B) is used for the selection of these synergies.
- A dynamic model is used to encode the succession in time of these synergies. This dynamic model specifies which synergy should be selected at time t knowing the synergy selected at time $t-1$ and the present internal state:

$$P(B^j|\pi_i) \times P(\beta_i^j|B^j B^{j-1} S_i^j \pi_i).$$

- The decision of a synergy to apply acts as a selection of the motor pattern to use:

$$P(M^j|\pi_i) \times P(\lambda_i^j|M^j S_i^j B^j M^{j-1}\pi_i)$$

The different steps to answer the ultimate question are:

- Predict the internal state corresponding to each area of interest (equation 6.4).
- Ask each area of interest for its recommendation about the appropriate synergy to select (equation 6.5).
- Merge the diverse opinions of the areas of interest in a coherent whole to select a synergy to apply (equation 6.6).
- Update the internal states of each area of interest, knowing which synergy has been selected (equation 6.7).
- Ask each area of interest for its motor recommendation knowing the applied synergy (equation 6.8).
- Finally, merge their opinion in a coherent whole to decide of the motor commands (equation 6.9).

We suggest that these local sensory-motor modules are possible mathematical implementation of the idea of motor synergies as proposed by Berstein [Bernstein, 1967]. Particularly in vertebrates, the body possesses too many degrees of freedom to be controlled by the brain. A relatively small number of coordinated movements (some very simple, others more complex) are chosen as natural movements and they are called synergies. This repertoire of coordinated movement constitutes the basis of all complex movements: a sequence of well adapted synergies are selected depending on the desired behaviour.

Indeed, each of the local sensory-motor modules describes a given sensory-motor behaviour adapted to a precise task (chasing a prey, for instance).

As seen in equations 6.21 and 6.27, the introduction of behaviour variable B does not lead to any notable reduction of either space complexity or time complexity of the ultimate question. However, introducing B greatly simplifies the intrinsic complexity of the motor models. In fact, B acts as a selector between the synergies and each of these synergies becomes much simpler and less complex to specify than the whole motor model.

At each time step, the decision of the adequate synergy using equation 6.6 produces a sequence of such synergies. We suggest that this process is a possible implementation of the idea of sensory-motor strategy as proposed by Berthoz [2000]. A strategy is then elaborated according to the sensory inputs, to the dynamic models describing the likely sequencing of the strategies and to the memory of what happened in the past synthesised by the probability distribution on the internal states, as can be seen in equation 6.5.

In this thesis, for efficiency reasons, we choose to make a decision on B to decide which synergy apply at each time step. It is also possible to avoid this decision and have B variable as a state variable, known only by a probability distribution. This alternative solution would imply to marginalise on variable B in all questions asked: the probability on B^t would be obtained as a weighted sum of the different dynamical models, where the weights are the estimated probability on the past internal states updated by sensor models. We think that this process could provide some mathematical insights to the discussion "switching versus weighing" and we propose that this matter should be investigated further.

Equation 6.7 shows a strong influence of synergy choice in the updating of state variables with sensory observations. A cognitive interpretation of this influence is that present behaviour biases the manner the world is interpreted: it may appear as a drawback, but it seems to have some psychological credibility.

6.4 Summary and Open Questions

This chapter has proposed a Behaviour Coordination mechanism to be included in the Bayes filter. Motor commands are grouped in patterns, called behaviours and fusion of elementary filter strategies for behaviour selection is employed to select which motor pattern to execute at the present time.

The joint distribution in each elementary filter now contains a behaviour selection modelling strategy for choosing a motor command pattern. The motor command model also has a dependence on behaviour when choosing the motor commands appropriate for the selected pattern:

$$\begin{aligned} & P(M^{0:t} S_i^{0:t} Z_i^{0:t} B^{0:t} \lambda_i^{0:t} \beta_i^{0:t} | \pi_i) \\ = & \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i). \end{array} \right. \end{aligned}$$

The prediction question in each elementary filter is unchanged:

$$\begin{aligned} & P(S_i^t | z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i) \\ \propto \; & \sum_{S_i^{t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(m^{t-1} | \pi_i) \times P(\lambda_i^{t-1} | m^{t-1} S_i^{t-1} b^{t-1} m^{t-2} \pi_i) \\ \times P(S^{t-1} | z_i^{0:t-1} b^{0:t-1} m^{0:t-2} \lambda_i^{0:t-2} \beta_i^{0:t-1} \pi_i) \end{array} \right] . \end{aligned}$$

Next, the behaviour selection question is asked of each elementary filter. Each elementary filter rates possible behaviours according to its own subset of state and observation variables.

$$
\begin{aligned}
& P(B^t | z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \pi_i) \\
& \propto \sum_{S_i^t} \left[\begin{array}{l} P(z_i^t | S_i^t \pi_i) \\ \times P(B^t | \pi_i) \times P(\beta_i^t | B^t b^{t-1} S_i^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i) \end{array} \right] .
\end{aligned}
$$

Combination of all the filter proposals is executed before choosing behaviour. The motor pattern chosen is a combination of all filter proposals.

$$
\begin{aligned}
& P(B^t | z^{0:t} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t} \pi_{bs}) \\
& \propto \prod_{i=1}^{N_i} P(B^t | z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \pi_i).
\end{aligned}
$$

Estimation then is calculated:

$$
\begin{aligned}
& P(S_i^t | z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \pi_i) \\
& \propto \left| \begin{array}{l} P(z_i^t | S_i^t \pi_i) \\ \times P(b^t | \pi_i) \times P(\beta_i^t | b^t b^{t-1} S_i^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i). \end{array} \right.
\end{aligned}
$$

According to the behaviour selected, the motor question is asked of each elementary filter. The answers express the collaboration of the elementary filters in the execution of the selected pattern.

$$
\begin{aligned}
& P(M^t | z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t} \beta_i^{0:t} \pi_i) \\
& \propto \sum_{S_i^t} \left[\begin{array}{l} P(M^t | \pi_i) \times P(\lambda_i^t | M^t S_i^t b^t m^{t-1} \pi_i) \\ \times P(S_i^t | z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \pi_i) \end{array} \right] .
\end{aligned}
$$

The combination of motor command answers is executed before drawing values for motor command variables:

$$
\begin{aligned}
& P(M^t | z^{0:t} b^{0:t} m^{0:t-1} \lambda^{0:t} \beta^{0:t} \pi_{bs}) \\
& \propto \prod_{i=1}^{N_i} P(M^t | z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t} \beta_i^{0:t} \pi_i).
\end{aligned}
$$

Once a motor pattern is selected, only some of the available observed environment features are used to control the robot's action.

Selecting a behaviour is directly related to controlling actuators, but it also concerns perception. The next chapter suggests employing an attention mechanism to reduce overhead in processing sensor data: only the most relevant features are chosen to be updated.

Chapter 7

Attention Selection

Key Points:

- **Focusing attention on the most relevant perceptual data reduces processing time.**
- **Attention is guided towards features pertinent to the next behaviour.**

Until now, no details have been given about how observations are acquired from the environment. All observations are assumed to be available at the moment when the behaviour selection question is asked.

Observations are defined as the results of processing sensor raw data. Sensors like vision often provide a huge amount of data, from which several features can be extracted by applying diverse processing algorithms. To have all observations for the Bayes filter, it can be necessary to execute different processing algorithms over sensor raw data, which is very time consuming.

In this chapter, we propose to reduce the overhead necessary to calculate all observations by applying the concept of attention, where only the most relevant perceptual observations are calculated and applied. Relevance of observation is defined as the features that are pertinent to the behaviour selected for robot execution.

7.1 Bayesian Program

The Bayesian program presented in this chapter includes a focus-of-attention model to decide which observations are more relevant at the present time step. This model contains a dependence on present behaviour.

Utilisation must now include an attention selection question in each elementary filter. The answers are combined, and a value is drawn from the resulting global distribution. This value indicates how to process the perceptions and, in consequence, which observations are assigned valid values.

A cyclic dependence then exists between selection of attention and selection of behaviour. To decide which data to process from perception, it would be desirable to know

which behaviour was most appropriate in the present situation. However, to decide which behaviour is most appropriate, it is necessary to have all available information from the environment. This cycle creates a problem regarding the order for asking the attention and behaviour questions.

The Bayesian approach provides a natural solution to this problem. The attention selection question is asked before the behaviour question, and when the inference rules are applied, the resulting expression indicates which behaviour is being predicted.

To have global behaviour prediction instead of local prediction in each elementary filter, the attention question in each elementary filter includes a behaviour set of variables, which are searched variables. The global attention question, however, concerns only attention selection variables.

7.1.1 Description

Additional Relevant Variables

To choose more relevant observations, a set of attention selection variables C^t is defined. Each variable in this set corresponds to one physical sensor or incoming flow of data. For one variable, each value indicates a different processing method to be applied to the raw data and the observation that results.

Attention variables C^t are common to all elementary filters, and the attention model is defined by applying coherence-based fusion. A coherence variable set α^t is then necessary, one variable in each elementary filter for each variable in C^t .

Decomposition

$$\begin{aligned} & P(M^{0:t} S_i^{0:t} Z_i^{0:t} B^{0:t} C^{0:t} \lambda_i^{0:t} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i) \\ & = \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j C^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\ \times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i). \end{array} \right. \end{aligned} \tag{7.1}$$

In the above joint distribution, the first term shows the dynamic model, which is unchanged from the Bayes filter in the previous chapter.

The observation model in the second line inside the time product now includes the attention variable:

$$P(Z_i^j | S_i^j C^j \pi_i).$$

The attention variable is necessary in the sensor model to indicate valid observations. If observations are valid, the sensor model applies. If an observation is not available, a uniform distribution is used. In other words, it is only possible to establish a

relation between sensor observations and state variables if these observations are available. Attention variables then control switching between a real sensor model and a uniform distribution.

The third line inside the product presents the behaviour selection model, which is unchanged from the previous version.

The next line introduces the attention selection model, where the attention variable depends on the present time states but also, and mainly, on present behaviour:

$$P(\alpha_i^t | C^t S_i^t B^t \pi_i).$$

The last line inside the time product shows the motor command model, unchanged from the previous chapter's Bayes filter, and the last line gives the initial conditions for all relevant variables.

Bayesian network representations of both elementary and global filter joint distributions are presented in appendix D, Figures D.1 and D.2, respectively.

7.1.2 Utilisation

Utilisation of this Bayes filter is similar to that in the previous chapter, with the addition of the attention selection question. This question is asked of each elementary filter after prediction: the answers are then combined, resulting in the global attention selection distribution; attention variables values are drawn from this global distribution.

The development of questions, derivation of expressions, and the proof of elementary filter recursive calculations are detailed in Appendix D.

Utilisation of the Bayes filter is illustrated in Figure 7.1. A new symbol (in orange, similar to a funnel) is employed in this figure to indicate that not all observations Z_i^t are available.

Prediction

$$\begin{aligned} & P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \\ \propto \quad & \sum_{S_i^{t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(m^{t-1} | \pi_i) \times P(\lambda_i^{t-1} | m^{t-1} b^{t-1} S_i^{t-1} m^{t-2} \pi_i) \\ \times P(S_i^{t-1} | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-2} \lambda_i^{0:t-2} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \end{array} \right]. \end{aligned} \tag{7.2}$$

Prediction in each elementary filter has not changed: in this expression, we can see the dynamic model, the influence of the motor command model and the drawn motor commands, and the estimation of states at the previous time step. Summation over the previous time step states is employed because the state variables for the previous time step do not have known values.

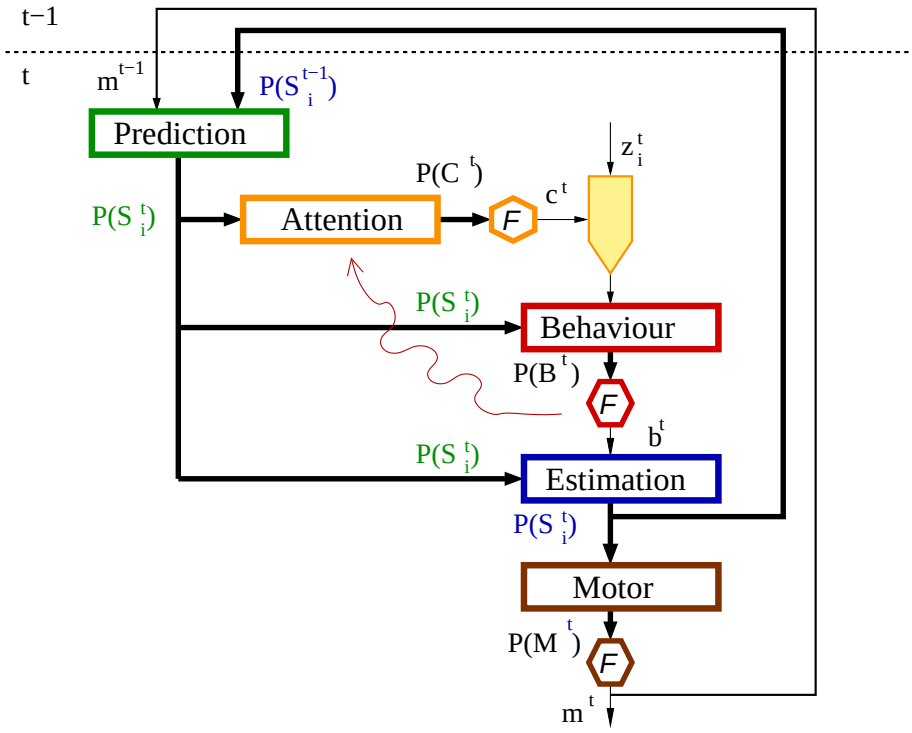


Figure 7.1: Utilisation of Bayes filter with behaviour coordination and attention selection. The red arrow indicates that the behaviour probability distribution is predicted at the moment of the attention question.

Attention

The attention selection question in the elementary filter is asked after the prediction question. Observations are not yet available, and they are not on the right side of this question.

The attention selection question asked in each elementary filter is then:

$$
\begin{aligned}
& P(C^t B^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} \pi_i) \\
& \propto \sum_{S_i^t} \left[\begin{array}{l} P(B^t|\pi_i) \times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\ \times P(C^t|\pi_i) \times P(\alpha_i^t | C^t S_i^t B^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \end{array} \right].
\end{aligned} \tag{7.3}
$$

In the left part of the question, it can be seen that not only attention variables C^t but also behaviour variables B^t are present. The attention model presents a dependence on the present-time behaviour variable, which is not yet known at the moment when the attention question is asked. To predict behaviour variable values, the behaviour variables are on the left side of the attention question. The coherence variables for attention (α_i^t) and behaviour (β_i^t) have values equal to one.

The expression answering this question shows the behaviour selection model in the first line inside the sum. The second line contains the attention selection model, and the last line contains the state prediction.

State variable values are not known, and to calculate the answer to the attention question, a summation over the state space is necessary to weight the behaviour and attention models according to the state prediction.

The answers to the attention question from the elementary filters are combined using the coherence-based fusion method, and the global attention expression is:

$$\begin{aligned} & P(C^t|z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as}) \\ & \quad \propto \ \sum_{B^t}\prod_{i=1}^{N_i}\{P(C^tB^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\alpha_i^{0:t}\pi_i)\}. \end{aligned} \tag{7.4}$$

Elementary filter answers to the attention question concern the attention and behaviour variables. At the moment of combining these proposals, only the attention variable is relevant. Consequently, it is desired to find the global distribution for the attention selection variables only.

To find this global distribution, the answers from the elementary filters are combined (by means of the distribution product) and a summation over behaviour variables is executed. This sum gives the attention variable global distribution results of weighting the attention model for all possible values of behaviour variables.

Once the global distribution over the attention selection variables is found, values for the attention variables are drawn, and the respective processing methods are applied to the sensor raw data. Corresponding observation variables are then assigned values resulting from the data processing, and valid observations are available.

Behaviour

Now that some observations are available, they can be used in the right part of the behaviour question, as well as the drawn values of attention variables C^t.

We assume that all variables Z^t have known values, even if only observations corresponding to values of C^t can be considered valid results of sensor data processing. The remaining observations are attributed non valid values, but they are not relevant during inference because of the uniform distributions in the sensor model.

The behaviour question asked to each elementary filter is then:

$$\begin{aligned} & P(B^t|z_i^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\alpha_i^{0:t}\pi_i) \\ & \quad \propto \ \sum_{S_i^t}\left[\begin{array}{l} P(z_i^t|S_i^tc^t\pi_i) \\ \times P(B^t|\pi_i)\times P(\beta_i^t|B^tS_i^tb^{t-1}\pi_i) \\ \times P(c^t|\pi_i)\times P(\alpha_i^t|c^tS_i^tB^t\pi_i) \\ \times P(S_i^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}\pi_i) \end{array}\right]. \end{aligned} \tag{7.5}$$

Inside the sum, the first line contains the sensor model. Its influence here is modulated by the known value of c^t: only available observations can be taken into account for updating states.

The second line shows the behaviour selection model, and the third line presents the attention model, with drawn values for attention variables. Its influence over behaviour variable values is similar to a valid observation and a sensor model.

The last line displays the state prediction. As state variables are not known, it is necessary to sum over the state space to weight the estimation of the state prediction distribution.

Fusion of the behaviour questions from each elementary filter results in a global distribution over the behaviour variables:

$$\begin{aligned} &P(B^t|z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as}) \\ &\quad\propto \prod_{i=1}^{N_i}\{P(B^t|z_i^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\alpha_i^{0:t}\pi_i)\}. \end{aligned} \tag{7.6}$$

From this global distribution, values for the behaviour variables are drawn.

Estimation

State estimation consists of updating the state prediction with all the new information available after prediction has taken place: drawn values for attention variables, observations, and drawn values for behaviour variables.

$$\begin{aligned} &P(S_i^t|z_i^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\alpha_i^{0:t}\pi_i) \\ &\propto \left| \begin{array}{l} P(z_i^t|S_i^tc^t\pi_i) \\ \times P(b^t|\pi_i)\times P(\beta_i^t|b^tS_i^tb^{t-1}\pi_i) \\ \times P(c^t|\pi_i)\times P(\alpha_i^t|c^tS_i^tb^t\pi_i) \\ \times P(S_i^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}\pi_i). \end{array} \right. \end{aligned} \tag{7.7}$$

In the first line, we can see the sensor model, where observations are related to states only for situations directly related to the drawn values of the attention variables.

The second line shows the behaviour selection model, with the drawn behaviour value. The third line contains the attention model, with the drawn value for the attention variable. The last line contains the state prediction.

The behaviour and attention models, as well as the drawn values for the behaviour and attention variables, propagate assumptions about the environment, influencing the estimation of state variables and their coherence with the selected values.

Motor

After estimating the state variables, the motor question is asked to each elementary filter:

$$\begin{aligned} & P(M^t|z_i^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t}\beta_i^{0:t}\alpha_i^{0:t}\pi_i) \\ & \propto \sum_{S_i^t}\left[\begin{array}{l} P(M^t|\pi_i)\times P(\lambda_i^t|M^tb^tS_i^tm^{t-1}\pi_i) \\ \times P(S_i^t|z_i^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\alpha_i^{0:t}\pi_i) \end{array}\right]. \end{aligned} \tag{7.8}$$

The expression for answering the motor question in the elementary filter has not changed from that in the previous chapter: the first line shows the motor model, with the drawn value for behaviour; the second line contains the state estimation; and as states are not known, a summation over state space is necessary to weight the motor command model in relation to the state estimation.

Fusion of the elementary filter propositions results in the global distribution for motor commands:

$$\begin{aligned} & P(M^t|z^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\alpha^{0:t}\pi_{as}) \\ & \propto \prod_{i=1}^{N_i}\{P(M^t|z_i^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t}\beta_i^{0:t}\alpha_i^{0:t}\pi_i)\}. \end{aligned} \tag{7.9}$$

Because coherence-based fusion is applied to define the motor command model, a simple product is sufficient to combine the elementary filters propositions.

7.1.3 Illustration Example

To illustrate the attention selection model in a Bayes filter, we return to our example, with the same behaviour specification as in the previous chapter.

Bayesian Program

We assume that our robot evolves in a real world, and it applies image processing to identify predator, prey and home in incoming image frames from a vision camera. A different image processing technique is used for identifying and localising each desired feature, according to colour, shape, motion pattern, and so on. Only one attention variable, C, is necessary, and each of its values indicates a different technique of image processing applied to the raw image.

We presented the home filter in Chapter 5 and the predator filter in Chapter 6. We now describe joint the distribution for the prey filter:

$$P(Y_{dir}^{0:t}Y_{dist}^{0:t}ZY_{dir}^{0:t}ZY_{dist}^{0:t}M_{rot}^{0:t}M_{trans}^{0:t}B^{0:t}C^{0:t}\lambda_{prey}^{0:t}\tau_{prey}^{0:t}\beta_{prey}^{0:t}\alpha_{prey}^{0:t}|\pi_{prey})$$

$$
= \left| \begin{array}{l} \Pi_{j=1}^{t} \left[\begin{array}{l} P(Y_{dist}^{j}|Y_{dist}^{j-1}\pi_{prey}) \times P(Y_{dir}^{j}|Y_{dir}^{j-1}M_{rot}^{j-1}\pi_{prey}) \\ \times P(ZY_{dir}^{j}|Y_{dir}^{j}C^{j}\pi_{prey}) \times P(ZY_{dist}^{j}|Y_{dist}^{j}C^{j}\pi_{prey}) \\ \times P(B^{j}|\pi_{prey}) \times P(\beta_{prey}^{j}|B^{j}B^{j-1}Y_{dist}^{j}\pi_{prey}) \\ \times P(C^{j}|\pi_{prey}) \times P(\alpha_{prey}^{j}|C^{j}B^{j}Y_{dist}^{j}\pi_{prey}) \\ \times P(M_{rot}^{j}|\pi_{prey}) \times P(\lambda_{prey}^{t}|M_{rot}^{j}Y_{dir}^{j}B^{j}\pi_{prey}) \\ \times P(M_{trans}^{j}|\pi_{prey}) \times P(\tau_{prey}^{t}|M_{trans}^{j}Y_{dist}^{j}B^{j}\pi_{prey}) \end{array} \right] \\ P(Y_{dir}^{0}Y_{dist}^{0}ZY_{dir}^{0}ZY_{dist}^{0}\lambda_{prey}^{0}\tau_{prey}^{0}M_{rot}^{0}M_{trans}^{0}B^{0}C^{0}\beta_{prey}^{0}\alpha_{prey}^{0}|\pi_{prey}). \end{array} \right. \tag{7.10}
$$

Inside the time product, the first line shows the dynamic model. That for the prey distance and direction is the same as described in equations 4.19 and 4.18.

Y_{dist}^{j}	0,1,2,3				4			
B^{j-1} / B^{j}	E	C	H	M	E	C	H	M
E	+	○	○	○	+	○	○	○
C	○	++	+	+	○	+	○	○
H	○	○	+	○	○	○	+	○
M	○	○	○	+	○	○	○	+

Table 7.1: Probability table for $P(B^{j}|B^{j-1}Y_{dist}^{j}\pi_{prey})$, behaviour selection inverse model for prey filter. E: escape behaviour; C: chase behaviour; H: homing behaviour; and M: motionless behaviour. ++ means highly recommended, + means recommended, and ○ means not relevant.

The second line presents the sensor model with attention modulation. The definition of the sensor model is very simple: if attention indicates prey, the sensor model is the same as in equations 4.24 and 4.25. A uniform distribution applies for all other values of the attention variable.

The behaviour selection model for the prey filter shown in the third line is the one employed in the previous chapter, but as it was not detailed there, Table 7.1 shows the inverse model.

Prey-related behaviour (chase) has a lower priority than those involved in defence against a predator, and there is no forbidden behaviour in this table. When prey is close and the robot is chasing, continuing to chase is highly recommended, but other behaviours are also possible, if other filters recommend them.

The fourth line in the joint distribution contains the attention model. Its inverse model is detailed in Table 7.2. It depends on the prey distance, but mostly on behaviour: if the robot is chasing, the recommendation to take information about the prey is the strongest.

Tables for the attention model in all filters are built in a similar manner, and the probability values are the same: ++ is given a value of 10 and + is 5; ○ is 1 and the value for – is 10^{-2}; $\times$ is attributed a value of 10^{-8}. Normalisation takes place after the table is filled and before utilisation for answering questions.

Y^j_{dist}	0,1,2				3,4			
C^j \ B^t	E	C	H	M	E	C	H	M
Predator	++	+	+	+	++	+	+	++
Prey	×	++	×	×	∘	++	∘	∘
Home	+	∘	++	∘	+	∘	++	∘

Table 7.2: Probability table for $P(C^j|Y^j_{dist}B^t\pi_{prey})$, the attention selection inverse model for the prey filter. E: escape behaviour; C: chase behaviour; H: homing behaviour; and M: motionless behaviour. ++ means highly recommended, + means recommended, ∘ means not relevant, and × means forbidden.

The motor command model in the fifth line of the joint distribution is defined as:

$$\begin{aligned} &P(M^j_{rot}|B^jY^j_{dir}\pi_{prey}) \\ &= \left\{ \begin{array}{l} P(M^j_{rot}|[B^t = Escape]\pi_{prey}); \\ P(M^j_{rot}|[B^t = Motionless]\pi_{prey}); \\ P(M^j_{rot}|Y^j_{dir}[B^t = Chase]\pi_{prey}); \\ P(M^j_{rot}|[B^t = Homing]\pi_{prey}). \end{array} \right. \end{aligned} \tag{7.11}$$

For chase behaviour, the rotation speed follows the prey direction:

$$P(M^j_{rot}|[B^t = Chase]Y^j_{dir}\pi_{prey}) = \mathcal{G}(M^j_{rot}), \mu = Y^j_{dir}, \sigma_{rot}. \tag{7.12}$$

As only chase behaviour is relevant in the prey filter, the other terms are uniform distributions.

The last line of the joint distribution contains the initial conditions.

The prediction question, as in the previous chapter, is asked to each elementary filter. In the prey filter, for example, we have the prediction question:

$$P(Y^t_{dir}Y^t_{dist}|zy^{0:t-1}_{dir}zy^{0:t-1}_{dist}c^{0:t-1}b^{0:t-1}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t-1}_{prey}\tau^{0:t-1}_{prey}\beta^{0:t-1}_{prey}\alpha^{0:t-1}_{prey}\pi_{prey}).$$

Following prediction, the attention question is asked of each elementary filter. For the prey filter, the attention question is:

$$P(C^tB^t|zy^{0:t-1}_{dir}zy^{0:t-1}_{dist}c^{0:t-1}b^{0:t-1}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t-1}_{prey}\tau^{0:t-1}_{prey}\beta^{0:t}_{prey}\alpha^{0:t}_{prey}\pi_{prey}).$$

The combination of attention answers from all elementary filters is given by:

$$\begin{aligned} &P(C^t|z^{0:t}_{lf}c^{0:t-1}b^{0:t-1}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t-1}_{prey}\tau^{0:t-1}_{prey}\beta^{0:t}\alpha^{0:t}\pi) \\ &\propto \sum_{B^t} \left[\begin{array}{l} P(C^tB^t|zx^{0:t}_{dir}zx^{0:t}_{dist}c^{0:t}b^{0:t-1}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t-1}_{pred}\tau^{0:t-1}_{pred}\beta^{0:t}_{pred}\alpha^{0:t}_{pred}\pi_{pred}) \\ \times P(C^tB^t|zy^{0:t}_{dir}zy^{0:t}_{dist}c^{0:t}b^{0:t-1}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t-1}_{prey}\tau^{0:t-1}_{prey}\beta^{0:t}_{prey}\alpha^{0:t}_{prey}\pi_{prey}) \\ \times P(C^tB^t|zh^{0:t}_{dir}zh^{0:t}_{dist}c^{0:t}b^{0:t-1}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t-1}_{home}\tau^{0:t-1}_{home}\beta^{0:t}_{home}\alpha^{0:t}_{home}\pi_{home}), \end{array} \right] \end{aligned}$$

where $z_{lf}^{0:t}$ is the set of all observation variables as defined in equation 5.15. Attention variable values are drawn from the above global distribution.

The behaviour question is then asked to each elementary filter, with the drawn attention value c^t included in the right part of the question. For the prey filter:

$$P(B^t|zy_{dir}^{0:t}zy_{dist}^{0:t}b^{0:t-1}c^{0:t}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{prey}^{0:t-1}\tau_{prey}^{0:t-1}\beta_{prey}^{0:t}\alpha_{prey}^{0:t}\pi_{prey}).$$

Next, the combination of answers of all elementary filters is calculated as:

$$\begin{array}{l} P(B^t|z_{lf}^{0:t}c^{0:t}b^{0:t-1}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{prey}^{0:t-1}\tau_{prey}^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi) \\ \propto \left| \begin{array}{l} P(B^t|zx_{dir}^{0:t}zx_{dist}^{0:t}c^{0:t}b^{0:t-1}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{pred}^{0:t-1}\tau_{pred}^{0:t-1}\beta_{pred}^{0:t}\alpha_{pred}^{0:t}\pi_{pred}) \\ \times P(B^t|zy_{dir}^{0:t}zy_{dist}^{0:t}c^{0:t}b^{0:t-1}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{prey}^{0:t-1}\tau_{prey}^{0:t-1}\beta_{prey}^{0:t}\alpha_{prey}^{0:t}\pi_{prey}) \\ \times P(B^t|zh_{dir}^{0:t}zh_{dist}^{0:t}c^{0:t}b^{0:t-1}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{home}^{0:t-1}\tau_{home}^{0:t-1}\beta_{home}^{0:t}\alpha_{home}^{0:t}\pi_{home}). \end{array} \right. \end{array}$$

From the above global distribution for the behaviour variable, a value for the present time behaviour variable is drawn.

The estimation question is then asked to each elementary filter, with the addition of the drawn value b^t in the right part of the question. For the prey filter, for example, we have:

$$P(Y_{dir}^tY_{dist}^t|zy_{dir}^{0:t}zy_{dist}^{0:t}c^{0:t}b^{0:t}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{prey}^{0:t-1}\tau_{prey}^{0:t-1}\beta_{prey}^{0:t}\alpha_{prey}^{0:t}\pi_{prey}).$$

After estimation, the motor question is asked of each elementary filter. For the prey filter, the motor question is:

$$P(M_{rot}^tM_{trans}^t|zy_{dir}^{0:t}zy_{dist}^{0:t}c^{0:t}b^{0:t}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{prey}^{0:t}\tau_{prey}^{0:t}\beta_{prey}^{0:t}\alpha_{prey}^{0:t}\pi_{prey}).$$

The combination of answers from all elementary filters is then:

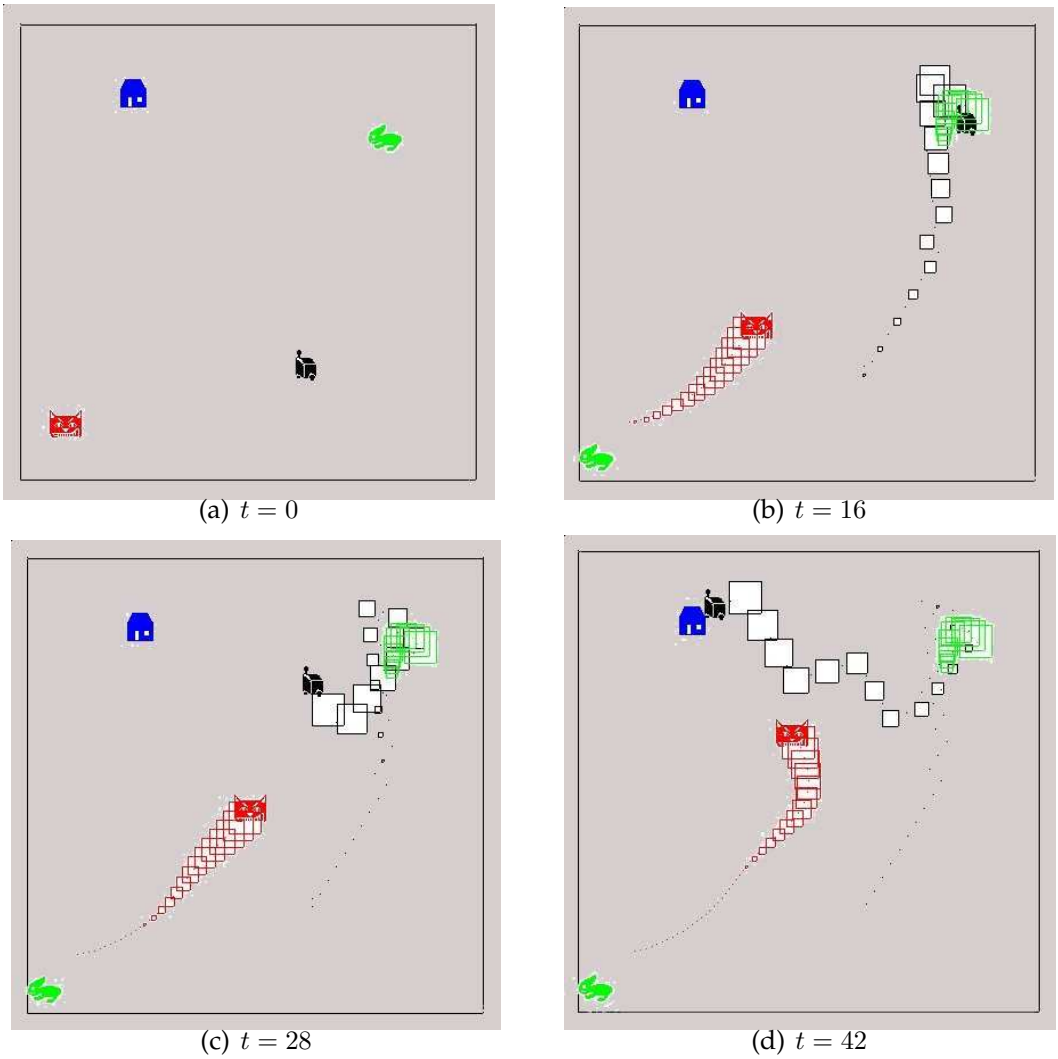
$$\begin{array}{l} P(M_{rot}^tM_{trans}^t|z_{lf}^{0:t}c^{0:t}b^{0:t}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{prey}^{0:t}\tau_{prey}^{0:t}\beta^{0:t}\alpha^{0:t}\pi) \\ \propto \left| \begin{array}{l} P(M_{rot}^tM_{trans}^t|zh_{dir}^{0:t}zh_{dist}^{0:t}c^{0:t}b^{0:t}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{home}^{0:t}\tau_{home}^{0:t}\beta_{home}^{0:t}\alpha_{home}^{0:t}\pi_{home}) \\ \times P(M_{rot}^tM_{trans}^t|zx_{dir}^{0:t}zx_{dist}^{0:t}c^{0:t}b^{0:t}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{pred}^{0:t}\tau_{pred}^{0:t}\beta_{pred}^{0:t}\alpha_{pred}^{0:t}\pi_{pred}) \\ \times P(M_{rot}^tM_{trans}^t|zy_{dir}^{0:t}zy_{dist}^{0:t}c^{0:t}b^{0:t}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{prey}^{0:t}\tau_{prey}^{0:t}\beta_{prey}^{0:t}\alpha_{prey}^{0:t}\pi_{prey}). \end{array} \right. \end{array}$$

Values of the motor command controlling the simulated robot are drawn from the above joint distribution.

Simulation Results

Snapshots of a simulation run are shown in Figure 7.2. Initial conditions are the same as for the simulation discussed in the previous chapter and the general behaviour is similar.

Nevertheless, prey capture shown in Figure 7.2(b), for example, occurs after a small deviation. The robot also escapes for some time before deciding to go home, as can be seen in Figure 7.2(c).



(a) $t = 0$ (b) $t = 16$

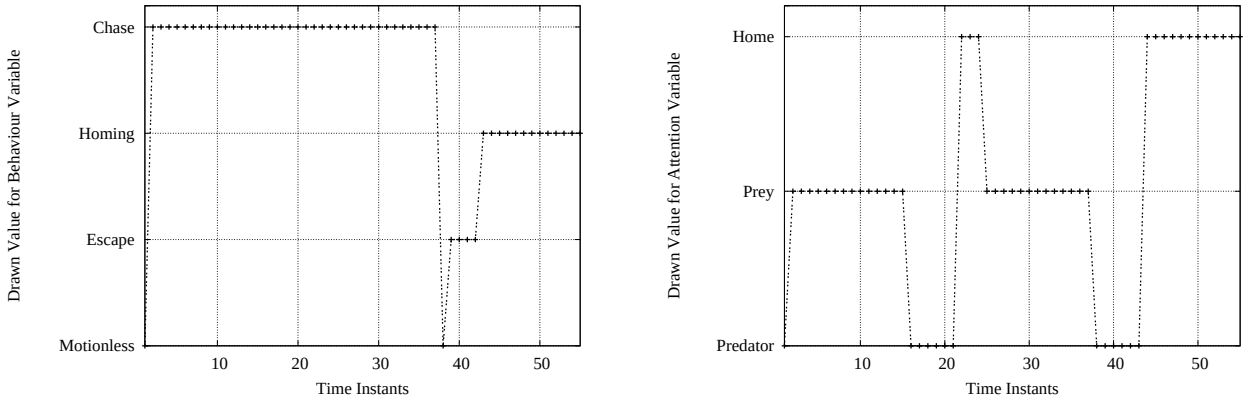
(c) $t = 28$ (d) $t = 42$

Figure 7.2: Sequence of snapshots for simulation of Example 1.

The drawn values for the behaviour and attention variables are shown in Figure 7.3. It is important to observe that attention selection oscillates during the prey hunt: from time $t = 0$ to time $t = 37$, the selected behaviour is always chase, but attention oscillates between predator, prey and home.

Two time instants illustrate very well the relation of attention selection to behaviour selection: at time instants $t = 16$ and $t = 38$, attention is switched from prey to predator. In the first situation, attention modulation does not cause a change in behaviour selection. In the second, changing attention is essential for provoking a behaviour transition. Figure 7.4 shows curves for the filter behaviour answers to behaviour questions, and for the resulting fusion.

Information acquired from the environment at time instant $t = 16$ is not relevant to changing behaviour selection, as can be seen in Figures 7.4(a) and 7.4(b): even though



(a) Drawn values for the behaviour variable

(b) Drawn values for the attention variable

Figure 7.3: Drawn values for the behaviour and attention variables for simulation of the illustrative example.

the attention variable has a different value, behaviour proposals from filters are not significantly altered by this change.

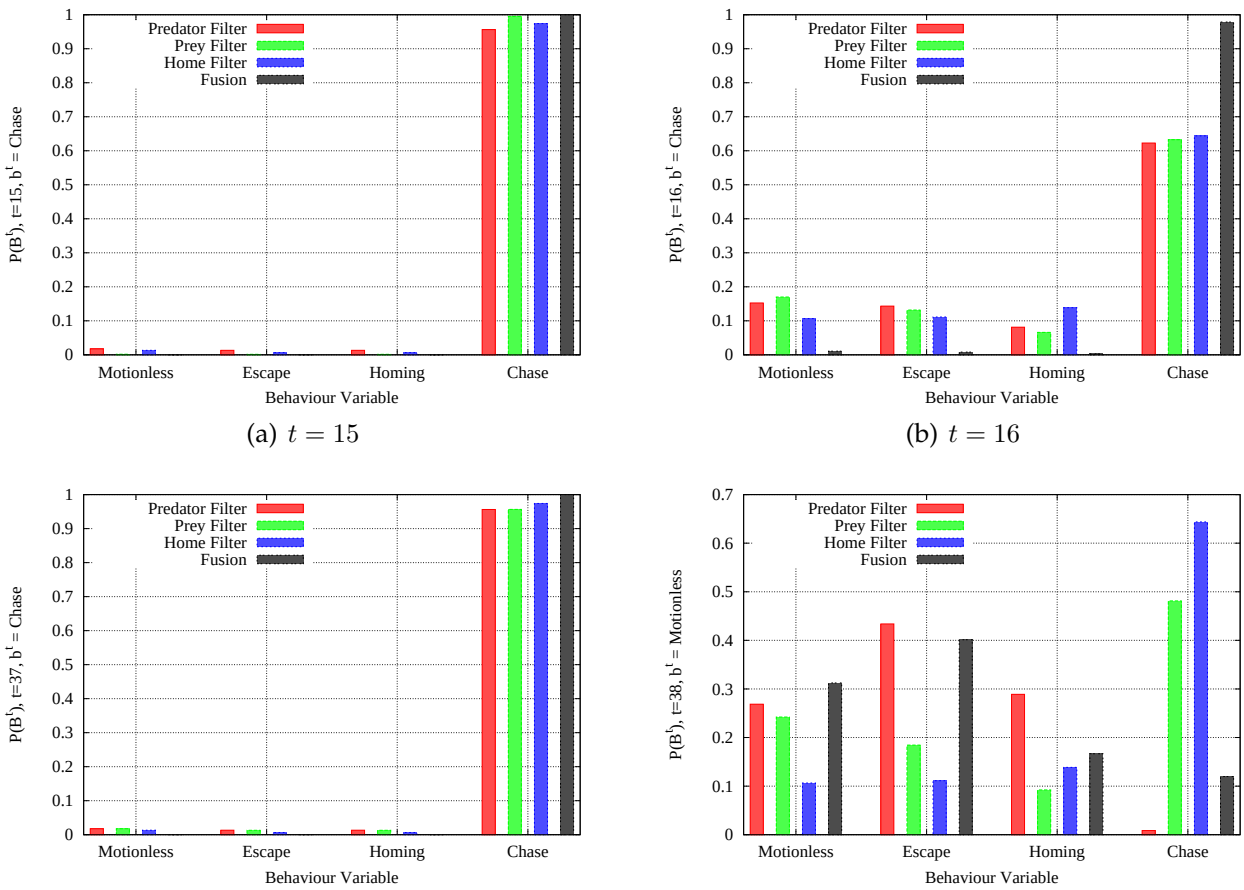
At time instant $t = 38$, however, a change in attention is responsible for modification of the behaviour curves, as shown in Figures 7.4(c) and 7.4(d). The main reason is that the observations updated were very pertinent for selecting behaviour.

The relevance of observations when estimating states can be observed in Figure 7.5. The curves show prediction (at left, in green) and estimation (at right, in blue) for the predator filter state variables. The top curves are related to time instant $t = 16$, and they show that, after the predator states are updated with new observations, the predator is considered far away (distance greater than three), and recommending defensive behaviour is not necessary.

On the other hand, at time instant $t = 38$, the predator state estimation indicates that the predator is close (distance three or less), as shown in Figure 7.5(d): the predator filter then suggests that selecting a defensive behaviour is important, as shown in Figure 7.4(d).

Figures 7.6 and 7.7 illustrate the influence of behaviour prediction on the selection of attention variables. The curves show, for each filter, the answers to the attention question and the resulting fusion. We can see that depending on behaviour, attention suggestion can be almost uniform, as shown by the Motionless and escape behaviours in the home and prey filters, for example. For some behaviours, the filter proposal for attention can have a bigger impact: chase behaviour in the home filter presents relevant information regarding the attention choice.

At time $t = 16$, all filters propose chase behaviour as more probable, judging by the curves 7.6(a), 7.6(b) and 7.6(c). Despite the fusion curve in Figure 7.6(d) being more favourable to choosing the prey value for the attention variable, the drawn value for attention is the predator value. As commented above, information acquired about the



(a) $t = 15$ (b) $t = 16$

(c) $t = 37$ (d) $t = 38$

Figure 7.4: Curves for filter propositions and the resulting fusion for the behaviour variable at different time instants.

predator does not change the behaviour proposed by the filter, which results in selecting chase behaviour.

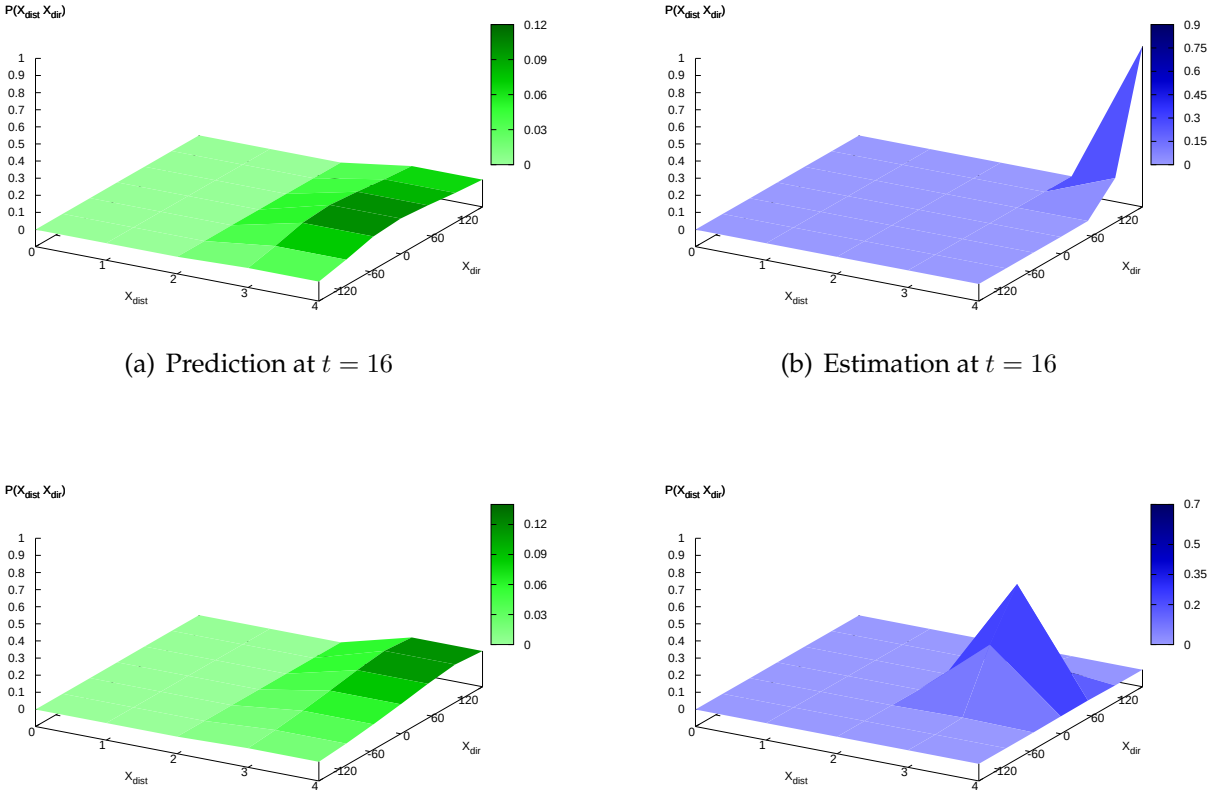
At time $t = 38$, the selected behaviour is Motionless, and at the following time instant, the predator filter answer for the attention question shows that defence behaviours are now more probable, as shown in Figure 7.7(a). The resulting fusion curve in Figure 7.7(d) shows that home and predator are more probable than prey, and the drawn value is predator. At time $t = 39$, the selected behaviour is Escape.

It is noticeable that the analysis of resulting behaviour becomes more complex. However, programming this behaviour using the proposed framework was rather simple.

7.1.4 Additional Main Assumptions

- Focus of attention reduces the time required for processing sensor data.

  In autonomous robotics, the environment provides a huge amount of sensor in-



(a) Prediction at $t = 16$

(b) Estimation at $t = 16$

(c) Prediction at $t = 38$

(d) Estimation at $t = 38$

Figure 7.5: Predictions and estimations for predator state variables in illustrative example simulation.

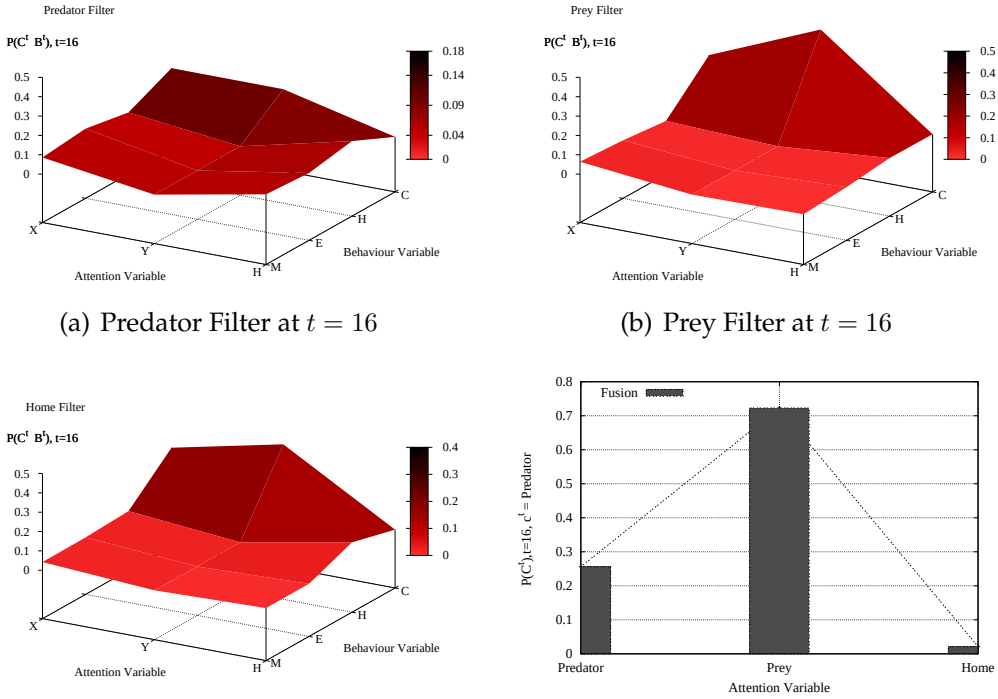
formation, which cannot be fully processed under real- time constraints. Focus of attention is a means for choosing a processing method according to the most relevant features desired from the environment, reducing the time spent treating sensor data.

- Attention aims to optimise the utility of the information available for the behaviour selected at the present time step.

  The most relevant features from the environment are defined here as the relevant ones for executing the selected behaviour. Attention selection must maximise the utility of observations in updating the most relevant features.

- Attention determines which observations are available for state updates.

  The sensor model defines how state variables affect observations. When attention is applied, observations are not fully available, and the sensor model cannot



(a) Predator Filter at $t = 16$

(b) Prey Filter at $t = 16$

(c) Home Filter at $t = 16$

(d) Fusion at $t = 16$

Figure 7.6: Attention Selection in elementary filters and the resulting fusion in the illustrative example simulation, at time instant $t = 16$.

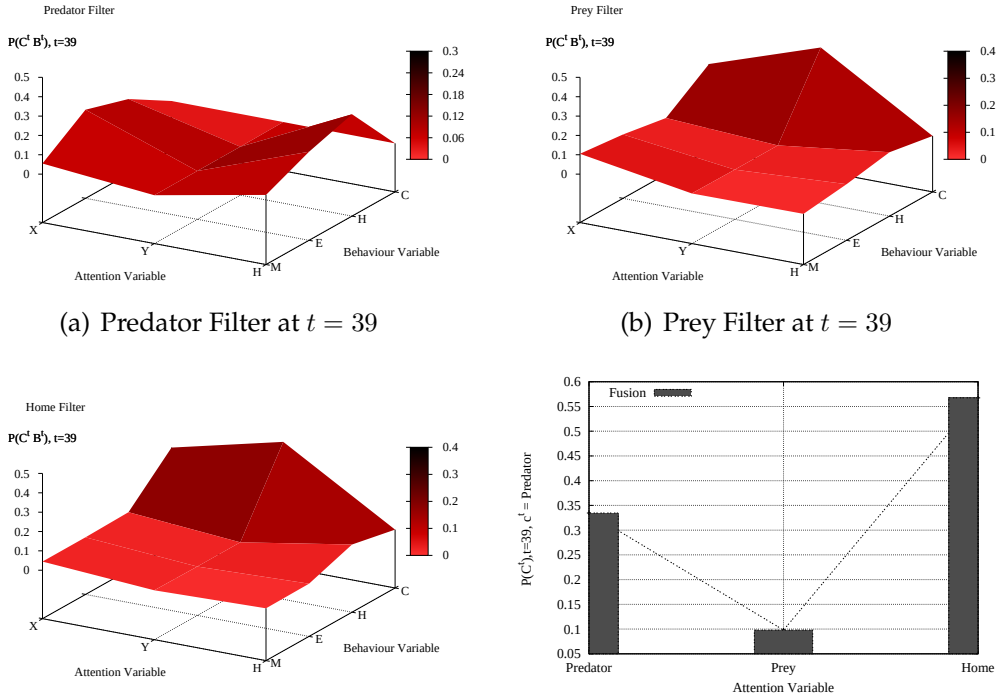
be employed unchanged. Attention is therefore added to the sensor model to indicate whether observations are part of the relevant features acquired from the environment.

7.2 Requirements Analysis

7.2.1 Space Complexity

The memory space required to store this model is calculated as the sum of storage for all terms in the joint distribution terms and the questions of each elementary filter. N_c stands for the number of attention variables, which are supposed to have the same number of cases n as other variables.

- A priori knowledge about states $P(S_i^{j-1}|\pi_i)$ requires $n^{N_{s_i}}$ cells.
- The dynamic model $P(S_i^j|S_i^{j-1}M^{j-1}\pi_i)$ requires $n^{2N_{s_i}+N_m}$ cells.
- The sensor model $P(Z^j|S_i^jC^j\pi_i)$ requires $n^{N_{z_i}+N_{s_i}+N_c}$ cells.



(a) Predator Filter at $t = 39$

(b) Prey Filter at $t = 39$

(c) Home Filter at $t = 39$

(d) Fusion at $t = 39$

Figure 7.7: Attention Selection in elementary filters and the resulting fusion in the illustrative example simulation, at time instant $t = 39$.

- The behaviour model $P(\beta_i^j | B^j S_i^j b^{j-1} \pi_i)$ requires $2n^{N_{s_i}+2N_b}$ cells.
- The attention model $P(\alpha_i^j | C^j S_i^j B^j \pi_i)$ requires $2n^{N_{s_i}+N_b+N_c}$ cells.
- The motor model $P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)$ requires $2n^{N_{s_i}+N_b+2N_m}$ cells.
- The prediction question requires $n^{N_{s_i}}$ cells.
- The local attention question requires $n^{N_b+N_c}$ cells.
- The local behaviour question requires n^{N_b} cells.
- The estimation question requires $n^{N_{s_i}}$ cells.
- The local motor question requires n^{N_m} cells.

Global questions for attention, behaviour and motor must also be stored. Summing all items, we have:

$$\mathcal{M} = \left| \begin{array}{l} \sum_{i=1}^{N_i} \left[\begin{array}{l} 3n^{N_{s_i}} + n^{2N_{s_i}+N_m} \\ +n^{N_{z_i}+N_{s_i}+N_c} + 2n^{N_{s_i}+2N_b} \\ +2n^{N_{s_i}+N_b+N_c} + 2n^{N_{s_i}+N_b+2N_m} \end{array} \right] \\ +(N_i+1)(n^{N_b} + n^{N_m} + n^{N_b+N_c}). \end{array} \right. \tag{7.13}$$

An upper bound for memory requirements is:

$$\mathcal{M} = O\left(\begin{array}{l} n^{2max(N_{s_i})+N_m} + n^{max(N_{z_i}+N_{s_i})+N_c} + n^{max(N_{s_i})+2N_b} \\ +n^{max(N_{s_i})+N_b+N_c} + n^{max(N_{s_i})+N_b+2N_m} + N_i(n^{N_b+N_c} + n^{N_m}) \end{array} \right). \tag{7.14}$$

Comparing this expression with equation 6.18:

$$\mathcal{M} = O\left(\begin{array}{l} n^{2max(N_{s_i})+N_m} + n^{max(N_{z_i}+N_{s_i})} + n^{2N_b+max(N_{s_i})} \\ +n^{2N_m+N_b+max(N_{s_i})} + N_i(n^{N_m} + n^{N_b}) \end{array} \right), \tag{7.15}$$

we can notice that a new term (in orange) was included in equation 7.14 for the attention model. The addition of attention variable dependencies in the sensor model also altered the blue term. A linear dependence on the number of elementary filters is present in the black term, with the addition of the number of attention variables. Again, the influence of state and observation variables is determined by the biggest subset.

For our illustrative example, the amount of required memory is 163 645 probability values. Comparing with previous chapter (161 393 probability values), the increment (2252 additional values to be stored) is rather small. Table 7.3 summarises the amount of memory space required for the examples in all chapters, including the naive table proposed in Chapter 1.

	Chapter 1	Chapter 4	Chapter 5	Chapter 6	Chapter 7
Space	10^{45}	10^{10}	111844	161393	163645
Time	???	10^{8}	11256	17078	21672

Table 7.3: Memory space and computational effort for illustrative examples in several chapters.

If sensor fusion is applied, as in previous chapters, the amount of required memory becomes:

$$\mathcal{M} = \left| \begin{array}{l} \sum_{i=1}^{N_i} \left[\begin{array}{l} 3n^{N_{s_i}} + n^{2N_{s_i}+N_m} \\ +N_{z_i} n^{N_{s_i}+N_c} + 2n^{N_{s_i}+2N_b} \\ +2n^{N_{s_i}+N_b+N_c} + 2n^{N_{s_i}+N_b+2N_m} \end{array} \right] \\ +(N_i+1)(n^{N_b} + n^{N_m} + n^{N_b+N_c}). \end{array} \right. \tag{7.16}$$

A new dependence appears in the sensor model, related to the number of attention variables. A new term related to the attention model is also added, and this term's influence on the amount of memory exceeds the influence of the sensor model.

Consequently, the amount of memory to store the Bayes filter proposed in this chapter has an upper bound of:

$$\mathcal{M} = O\left(\begin{array}{l} n^{2max(N_{s_i})+N_m} + max(N_{z_i})n^{max(N_{s_i})+N_c} + n^{max(N_{s_i})+2N_b} \\ +n^{max(N_{s_i})+N_b+N_c} + n^{max(N_{s_i})+N_b+2N_m} + N_i(n^{N_b+N_c} + n^{N_m}) \end{array} \right). \tag{7.17}$$

Comparing with equation 6.21 from the previous chapter:

$$\mathcal{M} = O\left(\begin{array}{l} n^{2max(N_{s_i})+N_m} + max(N_{z_i})n^{max(N_{s_i})} + n^{2N_b+max(N_{s_i})} \\ +n^{2N_m+N_b+max(N_{s_i})} + N_i \times (n^{N_m} + n^{N_b}) \end{array} \right), \tag{7.18}$$

we can notice the new term, in orange, related to the attention model, and the addition of the number of attention variables in the blue term, from the sensor model, and in the black term, related to the elementary filter answers to the attention question.

Let us assume the following relations between variables, true in most applications in robotics.

- The number of elementary filters is smaller than the number of state variables in the biggest filter:
$$N_i \ll n^{max(N_{s_i})}. \tag{7.19}$$
- The number of cases of all behaviour variables is smaller than the number of cases of all motor variables:
$$n^{N_b} \ll n^{N_m}. \tag{7.20}$$
- The number of cases of all behaviour variables is smaller than the number of cases of all state variables in the biggest filter:
$$n^{N_b} \ll n^{max(N_{s_i})}. \tag{7.21}$$
- The number of cases of all attention variables is smaller than the number of cases of all state variables in the biggest filter:
$$n^{N_c} \ll n^{max(N_{s_i})}. \tag{7.22}$$

The first three assumptions were presented and discussed in the previous chapter. The last one is related to the number of attention variables and comes from the role of the attention variables as model selectors: not all states are updated by the sensor models. This means that the number of cases of attention variables is smaller than both the number of observations and the number of state variables.

Considering the assumptions above, the space complexity expression becomes similar to the equations for space complexity in Chapter 5, as shown in the following equations: first for the case without sensor fusion, then when sensor fusion is applied.

$$\mathcal{M} = O\left(n^{2max(N_{s_i})+N_m} + n^{max(N_{z_i}+N_{s_i})} + n^{max(N_{s_i})+2N_m} \right). \tag{7.23}$$

$$\mathcal{M} = O\left(n^{2max(N_{s_i})+N_m} + max(N_{z_i})n^{max(N_{s_i})} + n^{2N_m+max(N_{s_i})} \right). \tag{7.24}$$

We conclude that adding the behaviour selection and attention selection models in most practical robotics applications does not imply much change in how memory requirements scale with system size.

7.2.2 Time Complexity

The necessary time for executing all calculations is based on the number of operations required for all the questions shown in Section 7.1.2. For each elementary filter, we have:

- The prediction question, as in equation 7.2, requires $4n^{2N_{s_i}} - n^{N_{s_i}}$ operations:

$$\begin{array}{l} P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \\ \propto \sum_{S_i^{t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(m^{t-1}|\pi_i) \times P(\lambda_i^{t-1} | m^{t-1} b^{t-1} S_i^{t-1} m^{t-2} \pi_i) \\ \times P(S_i^{t-1} | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-2} \lambda_i^{0:t-2} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \end{array} \right]. \end{array}$$

- The attention question, shown in equation 7.3, requires $4n^{N_{s_i}+N_b+N_c} - n^{N_b+N_c}$ operations:

$$\begin{array}{l} P(C^t B^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} \pi_i) \\ \propto \sum_{S_i^t} \left[\begin{array}{l} P(B^t|\pi_i) \times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\ \times P(C^t|\pi_i) \times P(\alpha_i^t | C^t S_i^t B^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \end{array} \right]. \end{array}$$

- The behaviour question defined in equation 7.5 requires $5n^{N_b+N_{s_i}} - n^{N_b}$ operations:

$$\begin{array}{l} P(B^t | z_i^{0:t} c^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} \pi_i) \\ \propto \sum_{S_i^t} \left[\begin{array}{l} P(z_i^t | S_i^t c^t \pi_i) \\ \times P(B^t|\pi_i) \times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\ \times P(c^t|\pi_i) \times P(\alpha_i^t | c^t S_i^t B^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \end{array} \right]. \end{array}$$

- The estimation question, as shown in equation 7.7, requires $4n^{N_{s_i}}$ operations:

$$\begin{array}{l} P(S_i^t | z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \pi_i) \\ \propto \left| \begin{array}{l} P(z_i^t | S_i^t \pi_i) \\ \times P(b^t|\pi_i) \times P(\beta_i^t | b^t b^{t-1} S_i^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i). \end{array} \right. \end{array}$$

- The motor question, as in equation 7.8, requires $3n^{N_{s_i}+N_m} - n^{N_m}$ operations:

$$\begin{aligned} & P(M^t|z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t}\beta_i^{0:t}\pi_i) \\ & \propto \sum_{S_i^t} \left[\begin{array}{l} P(M^t|\pi_i) \times P(\lambda_i^t|M^t S_i^t b^t m^{t-1}\pi_i) \\ \times P(S_i^t|z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\pi_i) \end{array} \right] . \end{aligned}$$

Additionally, fusions of questions are also required.

- Fusion of the attention question requires $N_i n^{N_b+N_c}$ operations:

$$\begin{aligned} & P(C^t|z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as}) \\ & \propto \sum_{B^t} \prod_{i=1}^{N_i} \{P(C^t B^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\alpha_i^{0:t}\pi_i)\}. \end{aligned}$$

- Fusion of the behaviour question requires $N_i n^{N_b}$ operation:

$$\begin{aligned} & P(B^t|z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as}) \\ & \propto \prod_{i=1}^{N_i} \{P(B^t|z_i^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\alpha_i^{0:t}\pi_i)\}. \end{aligned}$$

- Fusion of the motor question requires $N_i n^{N_m}$ operations:

$$\begin{aligned} & P(M^t|z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\pi_{bs}) \\ & \propto \prod_{i=1}^{N_i} P(M^t|z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t}\beta_i^{0:t}\pi_i). \end{aligned}$$

Summing all the above expressions, we have:

$$\mathcal{C} = \left| \sum_{i=1}^{N_i} \left[\begin{array}{l} 4n^{2N_{s_i}} + 3n^{N_{s_i}} \\ +4n^{N_c+N_{s_i}+N_b} \\ +5n^{N_b+N_{s_i}} \\ +3n^{N_{s_i}+N_m} \end{array} \right] \right. . \tag{7.25}$$

In the above equation, we can see an additional term compared with the corresponding equation in the previous chapter: this term is related to the attention question, and it also shows the dependence with behaviour variables.

For our illustrative example, the number of required operations at each time step is 21 672 compared with 17 078 in the previous chapter. Table 7.3 summarises the number of required operations for the same example in all proposed filters.

The number of operations to be executed has the upper bound:

$$\mathcal{C} = O\left(\begin{array}{l} n^{2max(N_{s_i})} + n^{max(N_{s_i})+N_m} \\ +n^{N_c+max(N_{s_i})+N_b} \end{array} \right). \tag{7.26}$$

Comparing with equation 6.27:

$$\mathcal{C} = O\left(n^{2max(N_{s_i})} + n^{N_m+max(N_{s_i})} + n^{N_b+max(N_{s_i})} \right), \tag{7.27}$$

we can see that the orange attention selection term in equation 7.26 replaces the behaviour selection term (shown in red in equation 6.27). This is because the attention question requires more effort than the answer to the behaviour question: in addition to summing over the state variables, a summation over the behaviour variables is also required. The green and brown terms in the above equation are unchanged from the previous chapter.

Applying the sensor fusion schema changes the requirement for operations to be executed, in both the behaviour and estimation questions:

$$\mathcal{C} = \sum_{i=1}^{N_i} \left[\begin{array}{l} 4n^{2N_{s_i}} + (4N_{z_i} - 1)n^{N_{s_i}} \\ +4n^{N_c+N_{s_i}+N_b} \\ +5N_{z_i} n^{N_b+N_{s_i}} \\ +3n^{N_{s_i}+N_m} \end{array} \right]. \tag{7.28}$$

$$\mathcal{C} = O\left(\begin{array}{l} n^{2max(N_{s_i})} + max(N_{z_i})n^{max(N_{s_i})} + max(N_{z_i})n^{N_b+max(N_{s_i})} \\ +n^{max(N_{s_i})+N_m} + n^{N_c+max(N_{s_i})+N_b} \end{array} \right). \tag{7.29}$$

Comparing with equation 6.30:

$$\mathcal{C} = O\left(\begin{array}{l} n^{2max(N_{s_i})} + max(N_{z_i})n^{max(N_s)} \\ +n^{N_m+max(N_{s_i})} + max(N_{z_i})n^{N_b+max(N_{s_i})} \end{array} \right), \tag{7.30}$$

a new term (in orange) is added, related to the cost of calculating the attention question.

For practical robotics applications, we can apply the same assumptions discussed in the space complexity analysis above, and we obtain the time complexity results in the following expressions (the first without sensor fusion, and second considering sensor fusion):

$$\mathcal{C} = O\left(n^{2max(N_{s_i})} + n^{max(N_{s_i})+N_m} \right). \tag{7.31}$$

$$\mathcal{C} = O\left(n^{2max(N_{s_i})} + max(N_{z_i})n^{max(N_{s_i})} + n^{max(N_{s_i})+N_m} \right). \tag{7.32}$$

The conclusion is that addition of the behaviour and attention questions does not change the way computational effort changes with system size.

7.2.3 Programming

The first step in programming the attention mechanism consists of choosing the attention variables and their semantics. Then, relations to the sensor model are defined, followed by writing the attention model.

Choice of Attention Variables

In the illustrative example proposed in this chapter, only one variable for attention selection is employed, and its semantics were related to the filters: actually, each value corresponded to updating the set of observation variables related to each elementary filter.

In a real robotics application, it can be interesting to associate one attention variable with each incoming flow of sensor data: vision, laser ranging, sonar data, and so on. All sensors will have their data processed but in different ways, and they may identify different features of the environment.

The main goal of attention selection is to reduce the time spent in processing data, so there can be no general formula. Any savings are specific to the specified behaviour, environment, and robot sensors. Development of a wider range of applications and more varied sensors will allow us to describe the most practical and effective configurations and semantics for attention variables.

Sensor Models

The next step in programming is to rewrite the sensor model using attention variables. The semantics associated with each attention variable are very important for this step, as is the definition of which observation variables correspond to each value of the attention variables.

The basic rule for rewriting the sensor model is: if the attention variables indicate that an observation is available, the sensor model agrees with the supplied sensor information. Otherwise, a uniform distribution is employed to mean that no information is available from this observation.

Attention Models

Last, it is necessary to write an attention model for each filter. Dependence on the present time behaviour and states is essential, but additional dependence on past-time attention, behaviour and motor commands is also possible without changing the calculation loop and fusion properties.

The easiest way to begin is to list the observations necessary for selecting and executing each possible behaviour: groups emerge from this list, with associated values of attention variables.

The inverse model for attention is responsible for relating current states and current selected behaviour with one of these groups of observations. It is desired to have no more than one computationally expensive observation related to each attention value.

When programming the behaviour model in Chapter 6, we suggested that the probability values should be classified as highly recommended, recommended, not recommended, unwise, and strictly prohibited. The same classification can be applied in the attention model, but unwise and prohibited situations must be used with care, otherwise corresponding observations may never (or only rarely) be updated.

An important consequence of applying an attention mechanism is that it greatly influences the selected behaviour. If attention selection is blocked and updates only one type of features, some features may not be updated for some time, and the robot behaviour can lose most of its reactivity.

7.3 Discussion

7.3.1 Robotics

Attention deployment has been the object of research and applications in artificial systems. Here, we address visual attention computational models and active/purposive vision, some of their application in robotics, and how they are related to our proposal.

Visual attention is defined as the brain's selective mechanism responsible for directing the gaze towards interesting objects in the visual environment [Itti, 2003a,b]. **Computational models of visual attention** are grouped according to the existence of intention, or voluntary deployment, of attention. Bottom-up control of visual attention is then based on image saliency features that attract attention, while top-down modulation of early vision is related to conscious and intentional selection of specific features on which to focus attention.

Most applications of visual attention computational models in robotics are related to gaze control, where a pan–tilt head is controlled to position one or several cameras.

A very interesting study was described by S.Vijayakumar et al. [2001]. It presented a computational model of attention implemented in a humanoid robot and it aimed to compare results obtained with this experiment with biological behaviour data. Some practical aspects of the implementation were particularly difficult, including real-time constraints, redundant degrees of freedom resulting from the head–eye combined movement, and imitation of peripheral and foveal vision. The robot task consisted of moving the head and eyes to fix an interesting location with its foveal vision; in other words, to execute attentional head–eye saccades.

Soyer et al. [2003] presented a mobile robot system showing properties similar to animal vision: retinal foveal–peripheral vision, attention, short-term memory, and others. The attention mechanism consisted basically of determining the position in the image where the fovea was focused, which is the most salient position according to the given function. Detailed visual processing was then executed only in the foveal image, which

is much smaller than the whole image. A sequence of images were employed for visual recognition of shapes in both simple and more complex scenes.

Another interesting and rather practical work was described by Baccon et al. [2002], where a mobile robot was guided by vision employing both bottom-up and top-down attention to execute tasks such as *approach an object*. Even if there is no explicit object recognition and the object has no internal complete representation, it was shown that the robot can approach a given object more easily if top-down bias is employed.

Active vision is based on the fact that all natural vision systems are active: to ease the recovery of information from the visual image, they control the image acquisition process (see Aloimonos [1993], for more details). Some ill-posed problems of general recovery in vision become well- posed by employing active vision.

A vision system is not required to process all available data all the time: that would require a huge amount of computational resources. To decide which processing method to apply, it is necessary to know which information is to be acquired by the vision, which depends on the relationship of the visual system with the world in which it lives, its task and its purpose.

Purposive vision is then an evolution of active vision, as it does not consider vision and perception as isolated, but as part of a more complex system situated in a world with which relations and interactions are established.

Rimey [1993] presented selective perception as purposive vision: scene analysis is selected based on immediate needs. Actually, Bayesian nets were applied to develop a system that decides how, where and when to activate a visual sensor and which method to apply to process its visual data. Simulation results were presented, showing the performance of the proposed system for simple and real-world scenes.

A focus-of-attention mechanism was proposed by Westelius [1995] to implement active vision in a robotic arm. Simulation results demonstrated that the foveal stereo vision model proposed for generating purposive actions for the robotic arm is fast and robust for gaze control of a stereo head.

Selective perception and short-term memory were both applied by McCallum [1995] to reduce or extend state spaces related to the perceptual system of an agent. When too much data is available, it is necessary to search for relevant information, and selective perception reduces the number of features considered. In the opposite situation, when relevant information is not included in the available data, short-term memory can be used to fill the gap. Reinforcement learning methods are applied to select features relevant to the current task, both in present and past (memory) sensor data.

Concerning application of active sensing in robotics, two major issues are: how to position sensors to acquire better information, and which trajectory to adopt to optimise gains and costs. Mihaylova et al. [2002] discussed this second point with emphasis on Bayesian solutions.

Active sensing in mobile robotics is mostly applied with laser range finder and proximity sensors, such as ultrasound and infrared. A typical task is localisation and mapping, as discussed by Fox et al. [1998a]. However, some studies also apply vision for mobile robot navigation [Davison, 1998].

Although coming from different scientific domains, with different aims and concerns, attention models of brain visual processing and active/purposive computer vision have several aspects in common. Above all, both methodologies were developed to deal with complexity in visual systems and optimise the utility of the acquired information.

Typical applications of these methodologies, as in the above mentioned systems, aim to model an attention system, or to add an attention model to a more complex behaviour system in simulation, or to verify attention mechanisms in simple robotic tasks. Active sensing is mostly employed for plan-based navigation, and sensors, including vision, are used to identify landmarks or environment features.

The influence of behaviour selection over perception was modelled by Tu [1996], where artificial animal models were applied to computer animation. Sensory data influences behaviour choice, and selected behaviour triggers the focuser, a perception module that drives the sensors according to intention. If the simulated male animal intention is to mate, the focus of attention is directed to locating a female.

In contrast, our approach aims to point out the importance of the reciprocal influence of attention selection and behaviour coordination. To decide correctly what to do, it is necessary to have new information about the robot surroundings; on the other hand, to select the most useful information, we must know what the robot is going to do.

We propose to solve this dilemma with the Bayes filter recursive loop of prediction and estimation. Prediction of behaviour allows us to decide on attention, and information acquired after focusing attention is used to estimate and draw a definite behaviour to be executed.

Utilisation of elementary filters helps to reduce computational effort, but it generates another problem: mathematically, it is possible to have behaviour prediction executed locally (in every filter) or globally (during attention fusion). This changes according to the left side of the attention question:

$$P(C^t B^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} \pi_i), \text{ or}$$

$$P(C^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} \pi_i).$$

There are two reasons for preferring the first form for the elementary filter attention question. The first is that the calculations are lighter, because only one summation over the behaviour variable space is required. The second reason is that attention results from a global prediction of behaviour, taking into account the combination of predictions from all elementary filters.

7.3.2 Cognition

The cognitive interpretations of the mathematical hypotheses added in this chapter (see decomposition in equation 7.1 are the following:

- The Central Nervous System does not allocate uniform effort to all the possible sensory data processing:
$$P(Z_i^j|S_i^j C^j \pi_i).$$
- It rather focuses its attention on some specific sensors according to its present internal state and to the present selected synergy:
$$P(C^j|\pi_i) \times P(\alpha_i^j|C^j S_i^j B^j \pi_i).$$

The different steps to answer the ultimate question are:

- Predict the internal state corresponding to each area of interest (equation 7.2).
- Ask each area of interest for its recommendation about where to focus attention according to next synergy (equation 7.3).
- Merge the diverse opinions of the areas of interest in a coherent whole to select the focus of attention (equation 7.4).
- Ask each area of interest for its recommendation about the appropriate synergy to select (equation 7.5).
- Merge the diverse opinions of the areas of interest in a coherent whole to select a synergy to apply (equation 7.6).
- Update the internal states of each area of interest, knowing where attention was focused and which synergy has been selected (equation 7.7).
- Ask each area of interest for its motor recommendation knowing where attention was focused and the applied synergy (equation 7.8).
- Finally, merge their opinion in a coherent whole to decide of the motor commands (equation 7.9).

Focusing attention mechanism does not reduce space and time complexity, as shown in equations 7.14 and 7.26. However, focusing attention on part of the sensory space and, consequently, neglecting other parts save a lot of cognitive efforts in signals processing.

Equation 7.7 shows that state variables outside the attention scope are maintained using the dynamic models, as in prediction phase. A global internal representation of the environment is then extrapolated from its known dynamic and continuity properties, as well as cumulated knowledge of its time evolution. Only part of this representation is updated with direct sensory inputs.

In our proposal, attention is modelled as an early decision in the behaviour selection process. Actually, the decision as to which features are to be perceived is based on anticipation of which behaviour will be selected. Equations 7.3 and 7.4 illustrates this

interesting property, as variable C is obtained as a weighted sum on all the possible values of B variable.

Vision begins in the eyes: an image is projected on the retina, and from there, specialised cells capture, convert and transfer information deep into the brain. Segregation of information at the lowest levels makes it possible to decrease the amount of information to be transmitted and processed by parts of the brain. The visual world is then analysed by decomposition in predefined categories: colour, form, motion. Recomposition is realised in relation to aims: advanced areas of visual processing can actually change the properties of the first layers to adapt to the present interest [Berthoz, 2002].

The first levels of perception (sensory data processing) are then adjusted according to this early attention selection, and perception is tuned to the stimulus necessary for the previewed behaviour. Incoming sensor data are understood according to the needs of the expected action.

7.4 Summary and Open Questions

This chapter proposed a focus-of-attention approach for dealing with rich environments and perception based on massive data from sensors. The main goal of this mechanism is to reduce the time spent processing the flow of sensor data: only the most relevant features for the selected behaviour are extracted from the incoming data.

An attention selection model is then added to the joint distribution of each elementary filter. Additionally, the sensor model now presents a new dependence, as the validity of observations is indicated by the attention variables. The joint distribution for each filter becomes:

$$
\begin{aligned}
&P(M^{0:t} S_i^{0:t} Z_i^{0:t} B^{0:t} C^{0:t} \lambda_i^{0:t} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i) \\
&= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j C^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\ \times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i). \end{array} \right.
\end{aligned}
$$

Utilisation begins with the prediction question, unchanged:

$$
\begin{aligned}
&P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \\
&\propto \sum_{S_i^{t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(m^{t-1} | \pi_i) \times P(\lambda_i^{t-1} | m^{t-1} b^{t-1} S_i^{t-1} m^{t-2} \pi_i) \\ \times P(S_i^{t-1} | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-2} \lambda_i^{0:t-2} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \end{array} \right].
\end{aligned}
$$

Before taking any observation into account, the attention selection question is asked of each elementary filter, including the behaviour variables. This is necessary to ensure a global behaviour prediction, instead of a local prediction in each elementary filter.

$$
\begin{aligned}
& P(C^t B^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} \pi_i) \\
& \quad \propto \sum_{S_i^t} \left[\begin{array}{l} P(B^t|\pi_i) \times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\ \times P(C^t|\pi_i) \times P(\alpha_i^t | C^t S_i^t B^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \end{array} \right].
\end{aligned}
$$

The combination of attention selection questions results in this expression, where the weighting over the behaviour variables indicates the prediction over the global distribution:

$$
\begin{aligned}
& P(C^t | z^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t} \alpha^{0:t} \pi_{as}) \\
& \quad \propto \sum_{B^t} \prod_{i=1}^{N_i} \{ P(C^t B^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} \pi_i) \}.
\end{aligned}
$$

From the global distributions over the attention variables, values for the attention variables are drawn, which indicate how to process the perception data.

After processing the sensor data, some observations are valid, and it is possible to ask the behaviour question of each elementary filter:

$$
\begin{aligned}
& P(B^t | z_i^{0:t} c^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} \pi_i) \\
& \quad \propto \sum_{S_i^t} \left[\begin{array}{l} P(z_i^t | S_i^t c^t \pi_i) \\ \times P(B^t|\pi_i) \times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\ \times P(c^t|\pi_i) \times P(\alpha_i^t | c^t S_i^t B^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \end{array} \right].
\end{aligned}
$$

Combining elementary filter answers yields the global distribution over the behaviour variables, from which behaviour variable values are drawn:

$$
\begin{aligned}
& P(B^t | z^{0:t} c^{0:t} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t} \alpha^{0:t} \pi_{as}) \\
& \quad \propto \prod_{i=1}^{N_i} \{ P(B^t | z_i^{0:t} c^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} \pi_i) \}.
\end{aligned}
$$

To this point, only simulation results were used to illustrate the proposals, but they are not suitable as evidence that the suggested framework can control a robot in a realistic environment with real- time constraints. The next chapter presents details and results of a practical implementation of the proposed programming framework on a real robot.

Chapter 8

Robot Experiments

Key Points:

- **Present a practical robotic implementation of the framework discussed the previous chapters.**
- **Present the Programming Methodology applied in this implementation.**
- **Analyse the benefits and disadvantages of the framework in a real application.**

In the previous chapters, we proposed a framework to control an autonomous sensory-motor system, but the results presented were mainly simulation examples employed as illustrations.

This chapter presents and discusses the results of practical utilisation of the previous chapters' proposals in a real robot. The implementation presented in this chapter was conducted in the context of an engineer training work and it is documented in [Pereira, 2005].

Initially, the experimental conditions are described in Section 8.1. Next, the Bayesian program is presented in Section 8.2, and the results obtained are presented and analysed in Section 8.3.

8.1 Experimental Setup

This section presents the experimental conditions. First, we describe the desired robot behaviour, as the specification for the experiment. Next, the robot itself is detailed: its physical structure, its interfaces with the external world and the on-board computational resources. Finally, we explain how the robot resources are used to implement the specified behaviour.

8.1.1 Behaviour Specification

In all previous chapters, simulation examples illustrating the proposal of each chapter used the same behaviour specification. We aim to have a similar behaviour implemented in a real robot.

Our robot wanders around in an indoor, office-like environment, avoiding obstacles. Whenever it perceives a predator, it stays motionless if the predator is far away, or it escapes in the opposite direction from the predator if it is close. When a prey is seen, the robot chases it, and when the prey is close enough, the robot captures it.

Homing behaviour, although simple to implement in a simulation, is in itself a very complex problem: it comprises navigation and landmark recognition actions, which are very interesting issues, but too demanding for this first verification.

Instead of implementing a complete homing behaviour, the concept of a trainer is included. It supplies additional security for the robot against predators, because when the robot is near the trainer, it is safe from predators. Basically, when the robot detects its trainer, it goes after him.

8.1.2 BIBA Robot Description

BIBA stands for *Bayesian-Inspired Brain and Artifacts* and it is the name of a European research project conducted from 2001 to 2005 (www-biba.inrialpes.fr). One of its objectives was to verify the utilisation of the Bayesian paradigm for programming artefacts. The BIBA Robot, shown in Figure 8.1, was conceived within this project as a tool for verifying theories and models, and this section presents its physical and logical structure.

The BIBA robot is an autonomous mobile robot that carries sensors (to perceive its surroundings), actuators (so that it can execute displacements or change the environment), and computation resources (to access sensors and actuators, execute calculations and communicate with other machines).

Among the on-board sensors we have a vision camera, a laser range finder, infrared and ultrasound proximity meters, and a vestibular sensor (measuring linear and angular speeds and accelerations). Only the vision camera and the laser range finder are used in this experiment.

As actuators, the BIBA robot has wheels for displacement and a pan–tilt head to change the direction of the vision camera sensor. In order to capture preys, a laser pointer is used, whose direction is controlled by the camera pan-tilt head. Another available actuator, not employed in this experimentation is a token placer, which can be used to simulate marks (pheromones, for example) that animals use to signal certain places.

In relation to computational resources, the BIBA robot has two embedded computers. One is responsible for controlling mechanical and electronic aspects of the robot, including real-time control of motors, sensors data reading, time stamp control, and

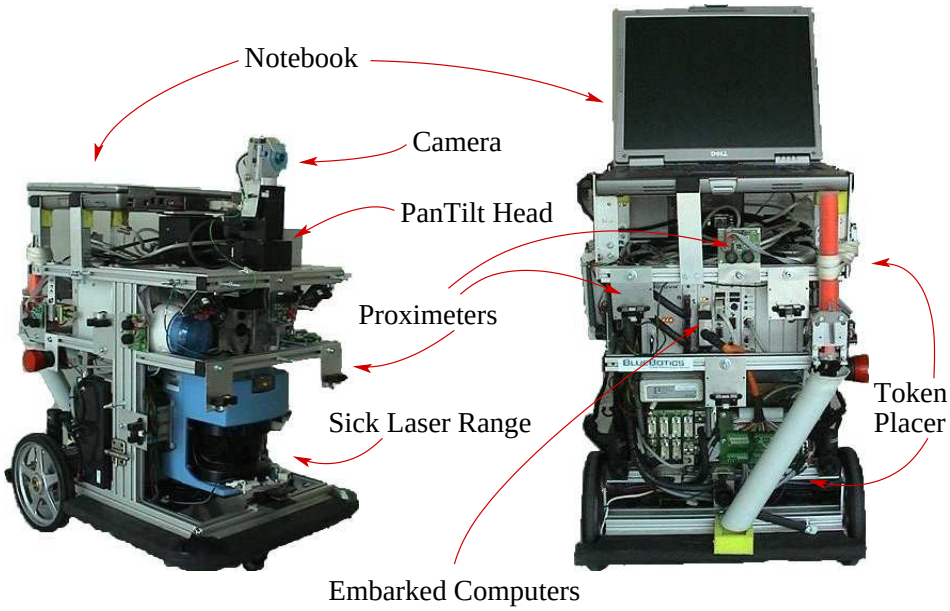


Figure 8.1: The BIBA robot: at left, a side view; at right, a back view.

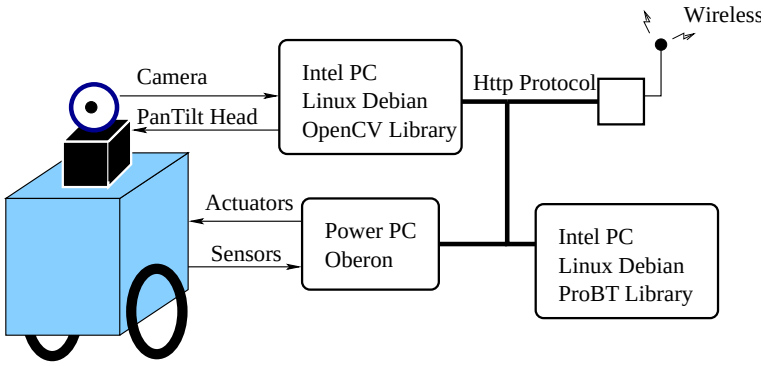


Figure 8.2: Robot BIBA and the computation structure employed in the experiments described in this chapter.

emergency stopping of wheel motors. The second computer is dedicated to computations at the level of executed task, which in our case are mostly probabilistic calculations.

The two embedded computers are linked in a local area network , also embedded in the robot. This local network has connections for two more computers, but only one is necessary in this experiment. This third computer (the notebook shown in Figure 8.1) is responsible for vision acquisition, image frame processing and pan– tilt head control. All machines communicate among themselves employing a socketed http protocol.

Figure 8.2 shows the computational structure and its interconnections. Operational systems and support libraries for execution of necessary tasks are indicated in each block in the diagram.

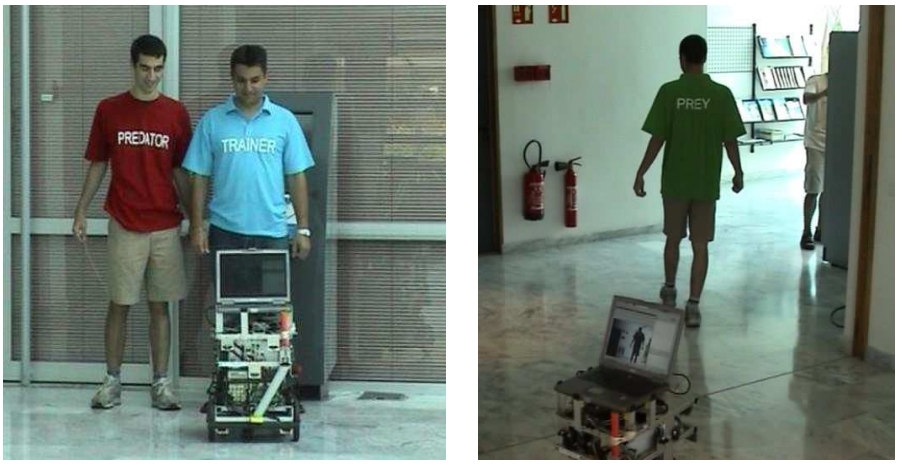


Figure 8.3: BIBA robot facing predator and trainer, at left. At right, the robot detects prey.

8.1.3 Behaviour Implementation in BIBA robot

When planning how to use robot resources to execute specified behaviours, we concentrate on how to interface the robot with its surroundings (via its sensors and actuators) and how to share the required computation burden of acquiring and processing data, and the calculations.

Sensors

Predator, prey and trainer are associated with colours and are detected by employing vision: the predator is red, the prey is green and the trainer is blue. It is assumed that humans will play these roles wearing coloured t-shirts. The left part of Figure 8.3 shows the robot facing the predator and the trainer, and in the right a prey is detected: the prey can be seen in the image displayed by the notebook.

The vision processing algorithm then identifies spots of these colours in image frames. Additional constraints in relation to minimal size and rough shape are applied to avoid detection of noise. More details of the vision processing can be found in Appendix G.

An escape route is detected by the laser range data processing: the robot's surroundings are explored to find the farthest direction, constrained to a minimum width. The farthest opening wider than 15 degrees is then considered to be the best available escape route.

Obstacles, for obstacle avoidance, are also identified using the laser range data. Section 8.2.3 discusses in detail the obstacle avoidance method employed in this experiment.

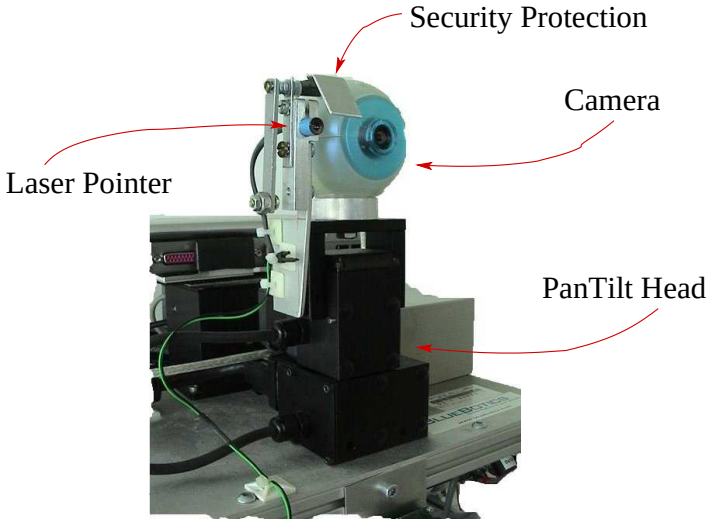


Figure 8.4: BIBA robot vision camera and laser capture actuator, with security cover.

Actuators

Some parts of the specified behaviour are difficult to achieve in a real robot. Consummation behaviour, for example, must be approximated, as the robot has no resources adapted for eating a prey.

Rather, a laser pointer is used to indicate when the robot is consuming a prey: when the robot is near enough, it flashes the prey with the laser pointer. Safety precautions are taken to avoid hurting people's eyes, as shown in Figure 8.4. An illustration of how capture takes place can be see in Figure 8.5, where the robot is pointing at a rather happy prey.

Computational Resources

The most complex computations consist of image processing and probabilistic calculations. The notebook is therefore in charge of the vision subsystem: the camera and pan–tilt head control, image acquisition and processing algorithm execution.

The pan–tilt head is embedded in the image processing, and it is controlled in terms of the spots found in the image. The biggest spot found is considered the most interesting, and the image is centred on this object.

Probabilistic calculations are performed in one of the on-board computers. It communicates with the vision computer (for acquiring observations resulting from image processing) and the controller computer (see Figure 8.2 for illustration), by means of a http protocol.

Calculations are programmed using the $ProBt^{©}$ library as inference engine. This library has been developed in eMotion research team, and it is commercialised by ProBayes company (www.probayes.com). Although this library includes several patent protected

Figure 8.5: BIBA robot capturing prey.

algorithms for reduction of calculations complexity, this implementation used exhaustive straightforward calculations.

8.2 Bayesian Programs

Bayesian programs are described in section 8.2.1 and utilisation follows the sequence described in section 8.2.2, with the addition of the obstacle avoidance stage. In fact, once motor commands values for present time step are known, it is important to assure their safe execution, which consists in assuring obstacle avoidance during execution of these motor commands. Description of filters utilisation including obstacle avoidance is presented in section 8.2.3.

8.2.1 Description

The Bayesian programs used in this experiment are described initially by listing all relevant variables. Then the global filter joint distribution, involving all relevant variables, is presented. The Conditional independence between state and observation variables defines four elementary filters, of which the trainer filter is described in details, followed by brief explanations of other filters.

Relevant Variables

In the simulation examples, the state variables in elementary filters were the distance and direction of the filter-related environment feature. Distance state variables are mostly used in behaviour and motor models as an indicator of feature presence: when near, feature is present; when far, feature can be considered absent.

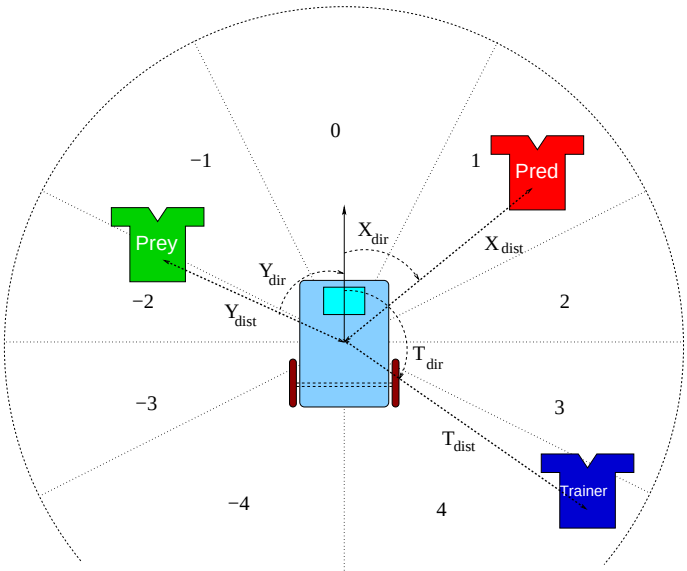


Figure 8.6: Robot forward direction and objects direction.

In this experiment, we explicitly add in each filter a state variable called presence, with two possible values (true and false). This state variable helps to simplify programming and simultaneously reduces computing time.

A distance variable here takes six possible values, and they are taken into account when the presence variable indicates that an object is present.

Direction follows the convention indicated in Figure 8.6:angles are measured clockwise from the robot's forward direction. Then, the value for direction is related to zones around robot: in the figure, the predator direction is 1, the prey direction is –2 and the trainer direction is 4. We can be seen that the zones behind the robot are larger (meaning a coarser indication of direction) than the zones in front of the robot. Main reason is that robot perception can detect objects in front of the robot only, and then determine their

direction. The zones behind the robot are employed for rough indication of probable feature position, and features located there are never perceived.

1. Common Variables to all Elementary Filters

   Motor commands are common to all filters and here they are rotation speed M_{rot} (nine values) and translation speed M_{trans} (six values). Figure 8.7 shows how translation and rotation speed are related to the robot.

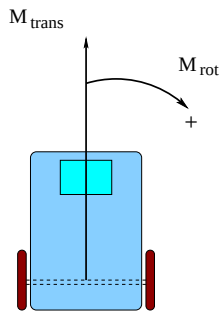


Figure 8.7: Robot rotation speed zones and Robot translation speed.

   Behaviour variable B^t is also common to all elementary filters, with values *escape*, *motionless*, *obey*, *chase* and *wander*. Escape, motionless and chase behaviour have the same meanings as explained in Example 2 of Section 4.1.3 and in Section 6.1.3.

   Obey behaviour means following the trainer closely, and wander behaviour denotes that the robot proceeds at a medium translation speed in the forward direction, avoiding obstacles.

   The head pan–tilt direction is an observation variable ZD_{pan} used by all filters, and it is employed to verify whether an object is in the robot's field of view.

2. Trainer Filter Relevant Variables

   State and observation variables are related to position of the trainer in relation to the robot: presence, direction and distance, with the types explained above.

$$S^j_{trai} = \{T^j_{pres} T^j_{dist} T^j_{dir}\}$$
$$Z^j_{trai} = \{ZT^j_{pres} ZT^j_{dist} ZT^j_{dist}\}$$

   Coherence variables are also necessary: one for each common variable. In this filter, we have λ^j_{trai}, for motor rotation speed, τ^j_{trai}, related to motor translation speed, and β^j_{trai}, for the behaviour variable.

3. Predator Filter Relevant Variables

   State and observation variables are related to position of the predator in relation to the robot: presence, direction and distance, with the types explained above.

$$S^j_{pred} = \{X^j_{pres} X^j_{dist} X^j_{dir}\}$$

When the predator is close, the robot escapes in the direction opposite to the predator. As a consequence, vision cannot supply information about the predator. In this specific situation, the dynamic model is not adequate to keep a record of predator presence long enough for the robot to escape. To provide a medium-term memory effect, an additional observation is employed by way of a timed presence variable ZX^j_{timer}.

$$Z^j_{pred} = \{ZX^j_{pres} ZX^j_{dist} ZX^j_{dir} ZX^j_{timer}\}$$

Coherence variables for this filter are λ^j_{pred}, τ^j_{pred}, and β^j_{pred}, for rotation speed, translation speed and behaviour variable, respectively.

4. Prey Filter Relevant Variables

   State and observation variables are related to position of the prey in relation to the robot: presence, direction and distance, with the types explained above.

$$S^j_{prey} = \{Y^j_{pres} Y^j_{dist} Y^j_{dir}\}$$
$$Z^j_{prey} = \{ZY^j_{pres} ZY^j_{dist} ZY^j_{dist}\}$$

   Coherence based fusion variables are: λ^j_{prey}, τ^j_{prey}, β^j_{prey}, for rotation speed, translation speed and behaviour variable, respectively. Motor command for capturing prey with laser pointer, M^j_{cons}, is relevant only in prey filter.

5. Escape Route Filter Relevant Variables

   State and observation variables are related to escape route position: as the distance of the escape route is not relevant, only its presence and direction are employed here.

$$S^j_{esc} = \{E^j_{pres} E^j_{dir}\}$$
$$Z^j_{esc} = \{ZE^j_{pres} ZE^j_{dist}\}$$

   Coherence based fusion variables are: λ^j_{esc}, τ^j_{esc} and β^j_{esc}, for rotation speed, translation speed and behaviour variable, respectively.

There are a total of fifty four variables, summing two hundred and seventy five cases for all variables. State variables are in number of twenty two variables and sum one hundred twenty four variables.

Decomposition

Global Joint Distribution

This joint distribution involves all variables listed above, and to simplify equations, the following set of variables is defined: the set of state variables, the set of observations variables, and the sets of coherence based fusion variables for behaviour, rotation speed and translation speed, respectively.

$$
\begin{aligned}
&P(S^{0:t} Z^{0:t} M_{rot}^{0:t} M_{trans}^{0:t} M_{cons}^{0:t} B^{0:t} \lambda^{0:t} \tau^{0:t} \beta^{0:t} | \pi) \\
&= \left| \prod_{j=1}^{t} \left[\begin{array}{l}
P(T_{pres}^{j} | T_{pres}^{j-1} \pi_{trai}) \times P(T_{dist}^{j} | T_{dist}^{j-1} B^{j-1} \pi_{trai}) \\
\times P(T_{dir}^{j} | T_{dir}^{j-1} B^{j-1} M_{rot}^{j-1} \pi_{trai}) \\
\times P(X_{pres}^{j} | X_{pres}^{j-1} \pi_{pred}) \times P(X_{dir}^{j} | X_{dir}^{j-1} B^{j-1} M_{rot}^{j-1} \pi_{pred}) \\
\times P(X_{dist}^{j} | X_{dist}^{j-1} B^{j-1} \pi_{pred}) \\
\times P(Y_{pres}^{j} | Y_{pres}^{j-1} \pi_{prey}) \times P(Y_{dir}^{j} | Y_{dir}^{j-1} B^{j-1} M_{rot}^{j-1} \pi_{prey}) \\
\times P(Y_{dist}^{j} | Y_{dist}^{j-1} B^{j-1} \pi_{prey}) \\
\times P(E_{pres}^{j} | E_{pres}^{j-1} \pi_{esc}) \times P(E_{dir}^{j} | E_{dir}^{j-1} B^{j-1} \pi_{esc}) \\
\times P(ZD_{pan}^{j} | \pi) \\
\times P(ZT_{dir}^{j} | T_{dir}^{j} ZD_{pan}^{j} ZT_{pres}^{j} \pi_{trai}) \\
\times P(ZT_{pres}^{j} | T_{pres}^{j} T_{dir}^{j} ZD_{pan}^{j} \pi_{trai}) \\
\times P(ZT_{dist}^{j} | T_{dist}^{j} T_{dir}^{j} ZD_{pan}^{j} ZT_{pres}^{j} \pi_{trai}) \\
\times P(ZX_{dir}^{j} | X_{dir}^{j} ZD_{pan}^{j} ZX_{pres}^{j} \pi_{pred}) \\
\times P(ZX_{pres}^{j} ZX_{timer}^{j} | X_{pres}^{j} X_{dir}^{j} ZD_{pan}^{j} \pi_{pred}) \\
\times P(ZX_{dist}^{j} | X_{dist}^{j} X_{dir}^{j} ZD_{pan}^{j} ZX_{pres}^{j} \pi_{pred}) \\
\times P(ZY_{pres}^{j} | Y_{pres}^{j} ZY_{dir}^{j} ZD_{pan}^{j} \pi_{prey}) \\
\times P(ZY_{dir}^{j} | Y_{pres}^{j} Y_{dir}^{j} ZD_{pan}^{j} \pi_{prey}) \\
\times P(ZY_{dist}^{j} | Y_{dist}^{j} Y_{dir}^{j} Y_{pres}^{j} ZD_{pan}^{j} \pi_{prey}) \\
\times P(ZE_{dir}^{j} | E_{dir}^{j} ZD_{pan}^{j} ZE_{pres}^{j} \pi_{esc}) \\
\times P(ZE_{pres}^{j} | E_{pres}^{j} ZE_{dir}^{j} ZD_{pan}^{j} \pi_{esc}) \\
\times P(B^{j} | \pi) \times P(\beta_{trai}^{j} | B^{j} B^{j-1} T_{pres}^{j} T_{dist}^{j} \pi_{trai}) \\
\times P(\beta_{pred}^{j} | B^{j} B^{j-1} X_{pres}^{j} X_{dist}^{j} \pi_{pred}) \\
\times P(\beta_{prey}^{j} | B^{j} B^{j-1} Y_{pres}^{j} Y_{dist}^{j} \pi_{prey}) \\
\times P(\beta_{esc}^{j} | B^{j} E_{pres}^{j} \pi_{esc}) \\
\times P(M_{rot}^{j} | \pi) \times P(M_{trans}^{j} | \pi) \\
\times P(\lambda_{trai}^{t} | M_{rot}^{j} B^{j} T_{dir}^{j} \pi_{trai}) \times P(\tau_{trai}^{t} | M_{trans}^{j} B^{j} T_{dist}^{j} \pi_{trai}) \\
\times P(\lambda_{pred}^{t} | M_{rot}^{j} B^{j} X_{dist}^{j} X_{dir}^{j} \pi_{pred}) \times P(\tau_{pred}^{t} | M_{trans}^{j} B^{j} X_{dist}^{j} \pi_{pred}) \\
\times P(\lambda_{prey}^{t} | M_{rot}^{j} B^{j} Y_{dir}^{j} \pi_{prey}) \times P(\tau_{prey}^{t} | M_{trans}^{j} B^{j} Y_{dist}^{j} \pi_{prey}) \\
\times P(M_{cons}^{j} | M_{cons}^{j-1} B^{j} Y_{pres}^{j} Y_{dist}^{j} Y_{dir}^{j} \pi_{prey}) \\
\times P(\lambda_{esc}^{t} | M_{rot}^{j} B^{j} E_{pres}^{j} E_{dir}^{j} \pi_{esc}) \times P(\tau_{esc}^{t} | M_{trans}^{j} B^{j} E_{pres}^{j} \pi_{esc})
\end{array} \right] \right. \\
&\left. \times P(S^{0} Z^{0} M_{rot}^{0} M_{trans}^{0} M_{cons}^{0} B^{0} \beta^{0} \lambda^{0} \tau^{0} | \pi) \right. .
\end{aligned}
\tag{8.1}
$$

$$
\begin{aligned}
S &= \{S_{trai} S_{pred} S_{prey} S_{esc}\} \\
Z &= \{Z_{trai} Z_{pred} Z_{prey} Z_{esc} ZD^{0:t}_{pan}\} \\
\beta &= \{\beta_{trai} \beta_{pred} \beta_{prey} \beta_{esc}\} \\
\lambda &= \{\lambda_{trai} \lambda_{pred} \lambda_{prey} \lambda_{esc}\} \\
\tau &= \{\tau_{trai} \tau_{pred} \tau_{prey} \tau_{esc}\}
\end{aligned}
$$

In this joint distribution, it is possible to see all the relevant variables and its relations. The colors correspond to the blocks in figure 1.2: the green terms are dynamic models; the sensor models are in blue; the red terms relate to behaviour coordination models; the brown terms are motor command models; and the initial conditions are in yellow.

Due to the conditioned independence between state and observation variables, it is possible to define four elementary filters, which simplifies the task of programming all the terms in the joint distribution.

Next section describes in details all the terms for the trainer filter. For the sake of brevity, the other three filters are not detailed, but only briefly discussed. Details about all filters are fully exposed in appendix H.

Trainer Filter

$$
\begin{aligned}
&P(S^{0:t}_{trai} Z^{0:t}_{trai} ZD^{0:t}_{pan} M^{0:t}_{rot} M^{0:t}_{trans} M^{0:t}_{cons} B^{0:t} \lambda^{0:t}_{trai} \tau^{0:t}_{trai} \beta^{0:t}_{trai} | \pi_{trai}) \\
&= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(T^j_{pres} | T^{j-1}_{pres} \pi_{trai}) \times P(T^j_{dist} | T^{j-1}_{dist} B^{j-1} \pi_{trai}) \\ \times P(T^j_{dir} | T^{j-1}_{dir} B^{j-1} M^{j-1}_{rot} \pi_{trai}) \\ \times P(ZD^j_{pan} | \pi_{trai}) \times P(ZT^j_{dir} | T^j_{dir} ZD^j_{pan} ZT^j_{pres} \pi_{trai}) \\ \times P(ZT^j_{pres} | T^j_{pres} T^j_{dir} ZD^j_{pan} \pi_{trai}) \\ \times P(ZT^j_{dist} | T^j_{dist} T^j_{dir} ZD^j_{pan} ZT^j_{pres} \pi_{trai}) \\ \times P(B^j | \pi_{trai}) \times P(\beta^j_{trai} | B^j B^{j-1} T^j_{pres} T^j_{dist} \pi_{trai}) \\ \times P(M^j_{rot} | \pi_{trai}) \times P(\lambda^j_{trai} | M^j_{rot} B^j T^j_{dir} \pi_{trai}) \\ \times P(M^j_{trans} | \pi_{trai}) \times P(\tau^j_{trai} | M^j_{trans} B^j T^j_{dist} \pi_{trai}) \end{array} \right] \\ \times P(S^0_{trai} Z^0_{trai} ZD^0_{pan} M^0_{rot} M^0_{trans} M^0_{cons} B^0 \lambda^0_{trai} \tau^0_{trai} \beta^0_{trai} | \pi_{trai}). \end{array} \right.
\end{aligned}
\tag{8.2}
$$

1. $P(T^j_{pres} | T^{j-1}_{pres} \pi_{trai})$

   The time dependence of the presence variable is shown in table 8.1. The first column shows the distribution over the trainer presence at time j knowing that the trainer was absent at time $j-1$: the trainer can arrive at any moment, but it is reasonable to assume that it is slightly more probable that the trainer is still absent.

   If the trainer was present at time $j-1$, probability of its presence at time j is shown in second column and it means that it is more probable that he is still near robot: this is because the robot tries to stay near the trainer.

T^j_{pres} \ T^{j-1}_{pres}	0	1
0	0.6	0.3
1	0.4	0.7

Table 8.1: Probability table for $P(T^j_{pres}|T^{j-1}_{pres}\pi_{trai})$.

2. $P(T^j_{dist}|T^{j-1}_{dist}B^{j-1}\pi_{trai})$

   A new dependence appears for the distance state variable compared with the simulation example. Here, the trainer distance dynamics depend on the previous time-step behaviour, while in the simulation example it was related to the previous time-step motor command.

   If robot present behaviour indicates that it is following the trainer, the distance is decreasing in time: this is not because the robot is faster than the trainer, but because it is assumed that the trainer is not escaping from robot and it desires to be followed. This assumption is expressed by table 8.2: most probable value for distance at time j is value of distance at time $j-1$ minus one.

T^j_{dist} \ T^{j-1}_{dist}	0	1	2	3	4	5
0	0.88889	0.5889	1e-1	1e-2	1e-3	1e-4
1	1e-1	0.3	0.489	1e-1	1e-2	1e-3
2	1e-2	1e-1	0.3	0.48	1e-1	1e-2
3	1e-3	1e-2	1e-1	0.3	0.489	1e-1
4	1e-4	1e-3	1e-2	1e-1	0.3	0.5889
5	1e-5	1e-4	1e-3	1e-2	1e-1	0.3

Table 8.2: Probability table for $P(T^j_{pres}|T^{j-1}_{pres}[B^{j-1} = Obey]\pi_{trai})$.

   For all other behaviours, it is not possible to say if trainer distance increases or decreases, and the dynamic model is defined as increasing uncertainty in time, by means of a bell-shaped distribution:

$$P(T^j_{dist}|T^{j-1}_{dist}[B^{j-1} \neq Obey]\pi_{trai}) = \mathcal{G}(T^j_{dist}), \mu = T^{j-1}_{dist}, \sigma = 0.5. \tag{8.3}$$

3. $P(T^j_{dir}|T^{j-1}_{dir}B^{j-1}M^{j-1}_{rot}\pi_{trai})$

   For the direction state variables, it is necessary to correct the direction in relation to the robot movement in the previous time step. If the robot is not following the trainer, a Gaussian curve is responsible for increasing the uncertainty about the new direction.

If the robot is following the trainer, the robot displacement in the previous time step was related to the trainer. The trainer direction is therefore corrected by multiplying the rotation speed in the previous time step by a constant K. This constant establishes the relation between rotation speed and trainer direction. It indicates the variation in the direction variable resulting from the rotation speed applied in the previous time step. The value of this constant is related to the duration of one time step and how the values of rotation speed are transformed into turns of the wheels and the consequent robot displacement. It is set empirically after realization of some practical experiments.

$$\begin{aligned} & P(T^j_{dir}|T^{j-1}_{dir}B^{j-1}M^{j-1}_{rot}\pi_{trai}) \\ & = \begin{cases} \mathcal{G}(T^j_{dir}), \mu = T^{j-1}_{dir}, \sigma = 0.5, \text{if } B^{j-1} \neq Obey; \\ \mathcal{G}(T^j_{dir}), \mu = T^{j-1}_{dir} + K * M^{j-1}_{rot}, \sigma = 0.5, \text{if } B^{j-1} = Obey. \end{cases} \end{aligned} \tag{8.4}$$

4. $P(ZD^j_{pan}|\pi_{trai})$

   The sensor model now includes the observation variable direction of pan–tilt head (ZD^j_{pan}). It is necessary here to identify whether the trainer is in the robot's field of view. The pan–tilt head is not controlled by this system but is fixed independently on the biggest object identified in the camera image: this is why the term $P(ZD^j_{pan}|\pi_{trai})$ shows no state dependence and is considered uniform.

5. $P(ZT^j_{pres}|T^j_{pres}T^j_{dir}ZD^j_{pan}\pi_{trai})$

   The presence variable can only be properly observed if the trainer is in the camera's field of view, i.e. located in the visible area. The visible area is determined by the observed head position and the trainer direction state variable, as illustrated in Figure 8.8. It can be seen that the predator direction is inside the visible area, but the trainer direction is outside. For this specific situation, observations related to the predator are taken into account in the predator filter, but observations regarding the trainer are not valid.

   For the trainer presence, the sensor model is then defined by Table 8.3, $|T^j_{dir} - ZD^j_{pan}| \leq 2$, that is, the trainer is inside the visible area. A uniform distribution is applied otherwise.

ZT^j_{pres} \ T^j_{pres}	0	1
0	0.7	0.1
1	0.3	0.9

Table 8.3: Probability table for $P(ZT^j_{pres}|T^j_{pres}T^j_{dir}ZD^j_{pan}\pi_{trai})$, if $|T^j_{dir} - ZD^j_{pan}| \leq 2$.

6. $P(ZT^j_{dist}|T^j_{dist}T^j_{dir}ZD^j_{pan}ZT^j_{pres}\pi_{trai})$

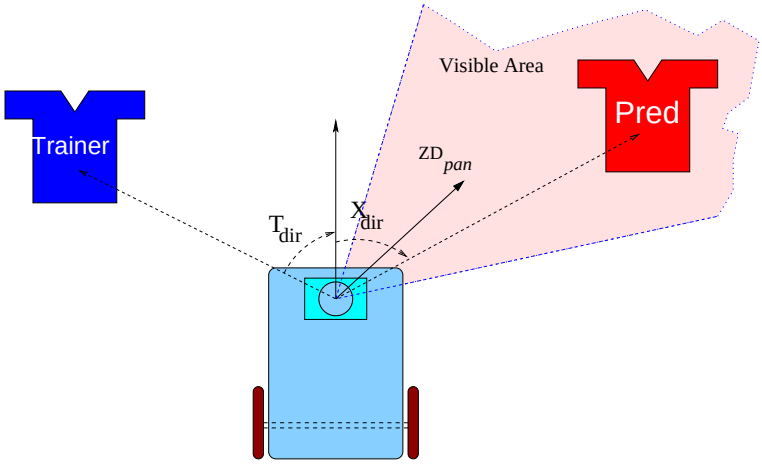


Figure 8.8: Illustration of robot visible area and head position.

The distance variable is updated by the distance observation only if the trainer is detected in the image ($ZT^j_{pres} = 1$) and in the camera's field of view ($|T^j_{dir} - ZD^j_{pan}| \leq 2$). Otherwise, a uniform distribution is applied.

$$P(ZT^j_{dist}|T^j_{dist}T^j_{dir}ZD^j_{pan}ZT^j_{pres}\pi_{trai}) = \begin{cases} \mathcal{G}(ZT^j_{dist}), \mu = T^j_{dist}, \sigma = 0.5, \text{if } ZT^j_{pres} = 1 \text{ and } |T^j_{dir} - ZD^j_{pan}| \leq 2; \\ \text{Uniform, otherwise.} \end{cases} \tag{8.5}$$

7. $P(ZT^j_{dir}|T^j_{dir}ZD^j_{pan}ZT^j_{pres}\pi_{trai})$

   Observation of the trainer direction is then determined by both the direction state and the head direction observation: if the trainer is detected in the image ($ZT^j_{pres} = 1$) and in the camera's field of view ($|T^j_{dir} - ZD^j_{pan}| \leq 2$).

$$P(ZT^j_{dir}|T^j_{dir}ZD^j_{pan}ZT^j_{pres}\pi_{trai}) = \begin{cases} \mathcal{G}(ZT^j_{dir}), \mu = T^j_{dir}, \sigma = 0.25, \text{if } ZT^j_{pres} = 1 \text{ and } |T^j_{dir} - ZD^j_{pan}| \leq 2; \\ \text{Uniform, otherwise.} \end{cases} \tag{8.6}$$

8. $P(\beta^j_{trai}|B^jB^{j-1}T^j_{pres}T^j_{dist}\pi_{trai})$

   The behaviour model expresses how behaviour can be selected based on information about the trainer. The behaviour inverse model is programmed as indicated in Table 8.4 and it is converted in Coherence based fusion form applying relations shown in equation 6.2.

   - Gray coloured columns of the table show the situation where the robot was escaping or motionless or chasing in the previous time step and the trainer is

B^{j-1}	E,M		C			O			W	
T^j_{pres}	0	1	0	1		0	1		0	1
T^j_{dist} / B^j	*	*	*	0..3	4..5	*	0..3	4..5	*	*
E	+	-	+	-	-	+	-	+	+	-
M	+	-	+	-	-	+	-	+	+	-
C	+	-	+	+	++	○	+	+	+	-
O	×	++	×	++	++	×	++	++	×	++
W	+	×	+	×	×	×	×	×	+	×

Table 8.4: Probability table for $P(B^j|B^{j-1}T^j_{pres}T^j_{dist}\pi_{trai})$, behaviour selection inverse model for the trainer filter. E: escape behaviour; C: chase behaviour; O: obey behaviour; and M: motionless behaviour. ++ means highly recommended, + means recommended, ○ means not relevant, – means unwise and × stands for forbidden. An asterisk indicates that the situation is the same for all values of one variable. The red cell can be read as *If the robot was chasing and the trainer is visible it is recommended that the robot continues to chase*.

not visible. As the trainer is absent, obey behaviour is forbidden and all other behaviours are considered recommended.

- In pale yellow columns, the robot was escaping or motionless or wandering in the previous time step and the trainer is visible now. When the trainer is present, the robot is safe from the predator, so behaviour obey is the highly recommended, no matter what the distance is to the trainer. Wander behaviour is forbidden, and all other behaviours are considered unwise.
- The robot was chasing in the previous time step and the trainer is visible now, at small distance: that is the situation in green column. If the trainer is visible and near, robot's safety is assured, and consequently the defensive behaviours (escaping and motionless) are considered unwise. Wandering behaviour is forbidden. Obey behaviour is highly recommended because the trainer is near, and chasing behaviour is still recommended.
- In pink column, the robot was chasing in the previous time step and the trainer is visible now, but far from the robot. If the trainer is visible and near, robot's safety is assured, and consequently the defensive behaviours (escaping and motionless) are considered unwise. Obey behaviour is highly recommended and wandering is forbidden. The trainer is far, so keep chasing is highly recommended.

  The differences between the green and the pink column are a practical example of the balance between persistence (keep chasing) and reactivity (follow trainer) in the robot global behaviour.
- White column presents the situation where the robot was following the trainer

in the previous time step and the trainer is not visible. In this case, obey behaviour is forbidden, as well as wandering. All other behaviours are equally recommended.

The Robot was following the trainer, and this could be a way to be in safety from a predator or even a distraction from a chasing behaviour. The state variables in the trainer filter do not provide any information about this, and it is not possible to determine which behaviour is the most interesting. Equal recommendation to possible behaviours is the method used to defer behaviour decision to the other filters.

- If the robot was following the trainer, and the trainer is visible and near, the most recommended behaviour is to obey, as shown in blue column. Wandering is forbidden, and defensive behaviours are unwise. Chasing is recommended, but this filter can not decide about it, as it has not the necessary informations about the presence and the distance of a prey.
- In orange column, the robot was following the trainer in the previous time step and the trainer is visible, but far. Defensive behaviours are recommended, and obey behaviour is highly recommended.

  Actually, the informations available in the Trainer filter are not enough to decide between defensive and obey behaviours: this depends a lot on how much defence is important. Consequently, this behaviour table makes possible to other filters to overcome the obey behaviour recommendation.
- Cyan column is related to the situation where the robot was wandering in the previous time step and the trainer is not visible now: obey behaviour is forbidden, and all other behaviours are recommended.

It is very important to stress that, when filling the behaviour table for the Trainer filter, the only informations taken into account are the ones available by means of filter specific subset of state variables: the trainer presence, distance and direction. This is the main rule of the proposed programming methodology for our framework, and it is applied in the definition of all joint distribution terms.

Whenever these informations are not enough, it is necessary to permit other filters to participate in the recommendation. This is possible by distributing probability values as not relevant or recommended among possible behaviours. The above reasoning is the basis of proscriptive programming and it is applied whenever the coherence-based fusion schema is used.

9. $P(\lambda^t_{trai}|M^j_{rot}B^jT^j_{dir}\pi_{trai})$

   The motor model consists of choosing a rotation speed and a translation speed for obey behaviour: for all others, a uniform distribution is applied. For executing the obey behaviour, the rotation speed is chosen to make robot turn in the trainer direction.

$$P(M^j_{rot}|B^jT^j_{dir}\pi_{trai}) = \begin{cases} \mathcal{G}(M^j_{rot}), \mu = T^j_{dir}, \sigma = 0.5, \text{if } B^j = Obey; \\ \mathcal{G}(M^j_{rot}), \mu = 0, \sigma = 8, \text{if } B^j = Wander; \\ \text{Uniform, otherwise.} \end{cases} \tag{8.7}$$

10. $P(\tau^t_{trai}|M^j_{trans}B^jT^j_{dist}\pi_{trai})$

    The translation speed in this filter is defined only if present time selected behaviour is obey. If the trainer is far, the robot translation speed is bigger; and if the trainer is near, the robot translation speed is smaller.

$$P(M^j_{trans}|B^jT^j_{dist}\pi_{trai}) = \begin{cases} \mathcal{G}(M^j_{trans}), \mu = T^j_{dist}, \sigma = 0.5, \text{if } B^j = Obey; \\ \mathcal{G}(M^j_{trans}), \mu = S_w, \sigma = 8, \text{if } B^j = Wander; \\ \text{Uniform, otherwise.} \end{cases} \tag{8.8}$$

Prey Filter

The prey filter is similar to the trainer filter with the addition of one motor variable, called M_{cons}, for activating the laser pointer. As this motor command concerns only the prey filter, it is not considered a common variable, and it is not defined using the coherence-based fusion schema.

The laser pointer is activated when the robot is chasing and prey is very close, ready to be consumed. To determine if prey is very close, it is necessary to verify if prey is present, if its distance to the robot is small, and if it is located in front of the robot. The motor model for the laser pointer command thus depends on present behaviour, as well as prey presence, distance and direction.

$$\begin{aligned} &P(M^j_{cons}|M^{j-1}_{cons}B^jY^j_{pres}Y^j_{dist}Y^j_{dir}\pi_{prey}) \\ &= \begin{cases} 1, \text{if } Y^j_{pres} = 1, Y^j_{dist} = 0 \text{ and } Y^j_{dir} = 0 \\ 0, \text{Otherwise} \end{cases} \end{aligned} \tag{8.9}$$

Predator Filter

The predator filter is also similar to the trainer filter, but a particularity is the addition of one sensor variable called ZX_{timer}. This observation is related to predator presence, and it becomes necessary when the robot executes an escape behaviour.

When the robot detects that the predator is present and close, it tries to escape in the opposite direction and should continue to escape for a certain time. However, the predator is no longer in the visible area, and observations regarding its presence will very likely result in false values (unless another predator is visible during the escape).

To strengthen the probability of predator presence in the absence of visual observations during execution of an escape behaviour, the observation ZX_{timer} is added in the

$X^j_{dir} ZD^j_{pan}$	$\lvert X^j_{dir} - ZD^j_{pan}\rvert \leq 2$		$\lvert X^j_{dir} - ZD^j_{pan}\rvert > 2$	
$ZX^j_{pres} ZX^j_{timer}$ \ X^j_{pres}	0	1	0	1
0 0	0.35	0.02	0.49	0.001
0 1	0.35	0.3	0.01	0.333
1 0	0.15	0.34	0.49	0.333
1 1	0.15	0.34	0.01	0.333

Table 8.5: Probability table for $P(ZX^j_{pres} ZX^j_{timer} | X^j_{pres} X^j_{dir} ZD^j_{pan} \pi_{pred})$, if $|X^j_{dir} - ZD^j_{pan}| \leq 2$.

predator filter. It works like a medium-term memory, preserving a true presence observation for some time. The sensor model for presence-related variables is then defined in Table 8.5.

If predator direction and head angle indicate that predator is within visible zone, but predator is in fact absent, visual observation is more probable to indicate false, as shown in the first column (grey colour). Timer observation is not relevant in this situation.

If predator is present inside visible area (second column, in yellow), it is unlikely that both observations are false; and it is equally probable that at least one observation indicates a present predator. A slight bigger probability is given for presence observation, because if it is inside visible area, it can be detected by image processing.

Timer observation variable has a very important role when predator is outside the visible area. If it not present (green column), timer variable indicates the predator is not there, and presence observation is not relevant. If predator is present it is possible that it cannot be seen). This is shown in the last column of table (in white) and it indicates that any true combination of both observations is probable: both indicating false is highly unlikely.

The behaviour model in the predator filter is also interesting. There is no specific behaviour associated with this filter: escape, motionless and even obey can be considered relevant behaviours when facing a predator. It is also important to note that this filter cannot decide by itself if it is better to escape (it has no information about the presence of an escape route) or if it is more interesting to stay near the trainer (as no information about the trainer is available). It is necessary to consider that posterior fusion of behaviour propositions for all filters can make a solution emerge because proposed by more than one filter, and not forbid behaviours that could be interesting in the present situation.

The behaviour model is programmed as shown in Table 8.6. It is evident that, whatever the selected behaviour was in the previous time step, when the predator is not present, wander, chase and obey behaviours are recommended, and escape and motionless are forbidden. This situation is indicated by the grey columns in table 8.6.

Green columns indicate the situation when the predator is present and the robot was already escaping: no matter what the predator distance is, continue to escape is

B^{j-1}	E		M			C			O,W		
X^j_{pres}	0	1	0	1		0	1		0	1	
B^j X^j_{dist}	*	*	*	0..3	4..5	*	0..3	4..5	*	0..3	4..5
E	×	++	×	++	+	×	++	+	×	++	+
M	×	-	×	-	++	×	-	++	×	-	++
C	+	-	+	-	-	+	-	+	+	-	-
O	+	+	+	+	+	+	+	++	+	+	++
W	+	×	+	×	×	+	×	×	+	×	×

Table 8.6: Probability table for $P(B^j|B^{j-1}X^j_{pres}X^j_{dist}\pi_{pred})$, behaviour selection inverse model for predator filter. E: Escape behaviour; C: chase behaviour; O: obey behaviour; and M: motionless behaviour. ++ means highly recommended, + means recommended, ∘ means unconcerned, – means unwise and × means forbidden. An asterisk indicates the situation is the same for all values of one variable.

most recommended, and obey is a good behaviour. Motionless and chase are unwise, and wander is forbidden. In this situation, the results from the other filters will add evidence: towards escape, by the escape route filter; or towards obey, if the trainer filter proposes it.

Cyan column shows that if robot was motionless in previous time step, and predator is present but far, it is highly recommended to stay motionless. Escape and obey are also recommended, but wandering is forbidden, and chasing is unwise.

When predator is present (but far) and robot was chasing in previous time step, obey and motionless behaviours are highly recommended. Escape is also recommended, as it is chasing, while wandering is forbidden. Further evidence from other filters influence the decision of which behaviour is most adequate. This situation is shown in yellow column of table 8.6.

A very similar situation happens in the last column (pink): robot was following trainer or wandering and now the predator is visible, but far. It is highly recommended that robot stays motionless or keep following trainer, as it is also recommended to escape. Continue to wander is forbidden, and chasing is unwise.

Escape Route Filter

The escape route filter is simpler than the trainer filter, especially in relation to the behaviour model, shown in Table 8.7. Actually, there is no recommendation for desired behaviour but an indication that escape is possible or not, according to the presence of a detected escape route.

8.2.2 Utilisation

Utilisation of the Bayesian program described above follows the diagram shown in figure 8.9, where questions to four elementary filters are indicated, as well as fusion of

B^j \ E^j_{pres}	0	1
E	×	+
M	○	○
C	○	○
O	○	○
W	○	○

Table 8.7: Probability table for $P(B^j|E^j_{pres}\pi_{esc})$, behaviour selection inverse model for escape route filter. E: escape behaviour; C: chase behaviour; H: homing behaviour; and M: for motionless behaviour. ++ means highly recommended, + means recommended, ○ means not relevant, – means unwise and × stands for forbidden. An asterisk indicates the situation is the same for all values of one variable.

behaviour and motor questions. Below, the sequence of asked questions is summarised and illustrated with expression solving questions to the trainer filter.

- The prediction question is asked of the four elementary filters.

$$P(S^t_{trai}|z^{0:t-1}_{trai}zd^{0:t-1}_{pan}b^{0:t-1}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t-1}_{trai}\tau^{0:t-1}_{trai}\beta^{0:t-1}_{trai}\pi_{trai})$$
$$\propto \sum_{S^{t-1}_{trai}} \left[\begin{array}{l} P(T^t_{pres}|T^{t-1}_{pres}\pi_{trai}) \times P(T^t_{dist}|T^{t-1}_{dist}b^{t-1}\pi_{trai}) \\ \times P(T^t_{dir}|T^{t-1}_{dir}b^{t-1}m^{t-1}_{rot}\pi_{trai}) \\ \times P(\lambda^{t-1}_{trai}|m^{t-1}_{rot}b^{t-1}T^{t-1}_{dir}\pi_{trai}) \\ \times P(\tau^{t-1}_{trai}|m^{t-1}_{trans}b^{t-1}T^{t-1}_{dist}\pi_{trai}) \\ \times P(S^{t-1}_{trai}|z^{0:t-1}_{trai}zd^{0:t-1}_{pan}b^{0:t-1}m^{0:t-2}_{rot}m^{0:t-2}_{trans}\lambda^{0:t-2}_{trai}\tau^{0:t-2}_{trai}\beta^{0:t-1}_{trai}\pi_{trai}) \end{array} \right] .$$

- Then, sensor data is acquired, processing is executed and observation variables values are assigned.

- The behaviour question is then asked of all elementary filters, employing the observation variable values.

$$P(B^t|z^{0:t}_{trai}zd^{0:t}_{pan}b^{0:t-1}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t-1}_{trai}\tau^{0:t-1}_{trai}\beta^{0:t}_{trai}\pi_{trai})$$
$$\propto \sum_{S^t_{trai}} \left[\begin{array}{l} P(zy^t_{dir}|T^t_{dir}zd^t_{pan}zt^t_{pres}\pi_{trai}) \\ \times P(zt^t_{pres}|T^t_{pres}T^t_{dir}zd^t_{pan}\pi_{trai}) \\ \times P(zt^t_{dist}|T^t_{dist}T^t_{dir}zd^t_{pan}zt^t_{pres}\pi_{trai}) \\ \times P(\beta^t_{trai}|B^tb^{t-1}T^t_{pres}T^t_{dist}\pi_{trai}) \\ \times P(S^t_{trai}|z^{0:t-1}_{trai}zd^{0:t-1}_{pan}b^{0:t-1}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t-1}_{trai}\tau^{0:t-1}_{trai}\beta^{0:t-1}_{trai}\pi_{trai}) \end{array} \right] .$$

- The answers to the behaviour questions are combined, and from the resulting global distribution, a value is drawn for the behaviour variable.

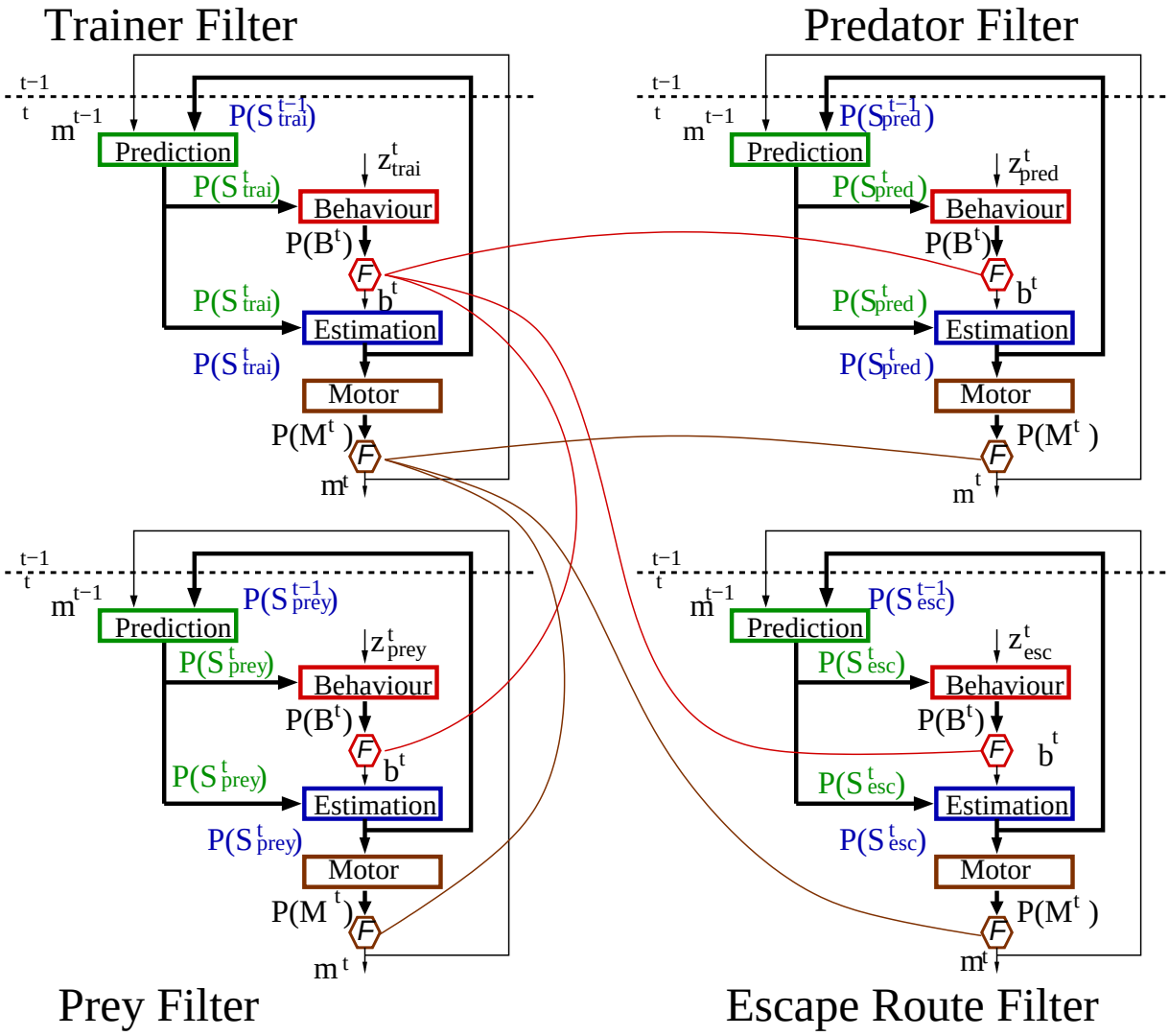


Figure 8.9: Utilisation of Bayes Filter for robot experiment.

$$
\begin{aligned}
&P(B^t|z^{0:t}b^{0:t-1}m_{rot}^{0:t-1}m_{trans}^{0:t-1}m_{cons}^{0:t-1}\lambda^{0:t-1}\tau^{0:t-1}\beta^{0:t}\pi)\\
&\propto \left|\begin{array}{l}
P(B^t|z_{trai}^{0:t}zd_{pan}^{0:t}b^{0:t-1}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{trai}^{0:t-1}\tau_{trai}^{0:t-1}\beta_{trai}^{0:t}\pi_{trai})\\
\times P(B^t|z_{pred}^{0:t}zd_{pan}^{0:t}b^{0:t-1}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{pred}^{0:t-1}\tau_{pred}^{0:t-1}\beta_{pred}^{0:t}\pi_{pred})\\
\times P(B^t|z_{prey}^{0:t}zd_{pan}^{0:t}b^{0:t-1}m_{rot}^{0:t-1}m_{trans}^{0:t-1}m_{cons}^{0:t-1}\lambda_{prey}^{0:t-1}\tau_{prey}^{0:t-1}\beta_{prey}^{0:t}\pi_{prey})\\
\times P(B^t|z_{esc}^{0:t}zd_{pan}^{0:t}b^{0:t-1}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{esc}^{0:t-1}\tau_{esc}^{0:t-1}\beta_{esc}^{0:t}\pi_{esc}).
\end{array}\right.
\end{aligned}
$$

- Next, the estimation question is asked of each elementary filter.

$$P(S_{trai}^t|z_{trai}^{0:t}zd_{pan}^{0:t}b^{0:t}m_{rot}^{0:t-1}m_{trans}^{0:t-1}\lambda_{trai}^{0:t-1}\tau_{trai}^{0:t-1}\beta_{trai}^{0:t}\pi_{trai})$$

$$\propto \left| \begin{array}{l} P(zy^t_{dir}|T^t_{dir}zd^t_{pan}zt^t_{pres}\pi_{trai}) \\ \times P(zt^t_{pres}|T^t_{pres}T^t_{dir}zd^t_{pan}\pi_{trai}) \\ \times P(zt^t_{dist}|T^t_{dist}T^t_{dir}zd^t_{pan}zt^t_{pres}\pi_{trai}) \\ \times P(\beta^t_{trai}|b^t b^{t-1}T^t_{pres}T^t_{dist}\pi_{trai}) \\ \times P(S^t_{trai}|z^{0:t-1}_{trai}zd^{0:t-1}_{pan}b^{0:t-1}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t-1}_{trai}\tau^{0:t-1}_{trai}\beta^{0:t-1}_{trai}\pi_{trai}). \end{array} \right.$$

- The motor question is asked of all four filters.

$$P(M^t_{rot}M^t_{trans}|z^{0:t}_{trai}zd^{0:t}_{pan}b^{0:t}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t}_{trai}\tau^{0:t}_{trai}\beta^{0:t}_{trai}\pi_{trai})$$
$$\propto \sum_{S^t_{trai}} \left[\begin{array}{l} P(\lambda^t_{trai}|M^t_{rot}b^tT^t_{dir}\pi_{trai}) \\ \times P(\tau^t_{trai}|M^t_{trans}b^tT^t_{dist}\pi_{trai}) \\ \times P(S^t_{trai}|z^{0:t}_{trai}zd^{0:t}_{pan}b^{0:t}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t-1}_{trai}\tau^{0:t-1}_{trai}\beta^{0:t}_{trai}\pi_{trai}) \end{array} \right].$$

- The answers to the motor questions are combined and from the resulting global distribution, values are drawn for the motor command variables.

$$P(M^t_{rot}M^t_{trans}M^t_{cons}|z^{0:t}b^{0:t}m^{0:t-1}_{rot}m^{0:t-1}_{trans}m^{0:t-1}_{cons}\lambda^{0:t}\beta^{0:t}\pi_{bs})$$
$$\propto \left| \begin{array}{l} P(M^t_{rot}M^t_{trans}|z^{0:t}_{trai}zd^{0:t}_{pan}b^{0:t}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t}_{trai}\tau^{0:t}_{trai}\beta^{0:t}_{trai}\pi_{trai}) \\ \times P(M^t_{rot}M^t_{trans}|z^{0:t}_{pred}zd^{0:t}_{pan}b^{0:t}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t}_{pred}\tau^{0:t}_{pred}\beta^{0:t}_{pred}\pi_{pred}) \\ \times P(M^t_{rot}M^t_{trans}M^t_{cons}|z^{0:t}_{prey}zd^{0:t}_{pan}b^{0:t}m^{0:t-1}_{rot}m^{0:t-1}_{trans}m^{0:t-1}_{cons}\lambda^{0:t}_{prey}\tau^{0:t}_{prey}\beta^{0:t}_{prey}\pi_{prey}) \\ \times P(M^t_{rot}M^t_{trans}|z^{0:t}_{esc}zd^{0:t}_{pan}b^{0:t}m^{0:t-1}_{rot}m^{0:t-1}_{trans}\lambda^{0:t}_{esc}\tau^{0:t}_{esc}\beta^{0:t}_{esc}\pi_{esc}). \end{array} \right.$$

- Estimation is used to update the a priori term in the joint distribution for utilisation in the next time step.

After drawing values for motor commands, they must be executed safely, avoiding obstacles that may appear in the robot's path.

Obstacle avoidance is independent of the Bayes filter: it receives the desired motor commands as set points and evaluates whether their execution is compatible with the detected positions of obstacles. The resulting motor commands are then sent to the controller to be executed in the robot actuators.

The Bayesian program used to implement obstacle avoidance is described in the following section.

8.2.3 Safe Execution of Motor Commands

The Bayesian program described above is used to infer motor command to be applied to the robot actuators in order to execute the specified behaviour. But it is important to assure that these commands can be safely executed: the robot is not going to touch any obstacles while executing them.

We assume an upper level module supplies references of commands to be followed by the robot. These reference commands are chosen according to the desired behaviour and are called M_{rot}, for rotation speed reference variable, and M_{trans}, for translation speed reference variable.

A lower level module is responsible for providing commands that are going to be applied in robot actuators (V_{trans}, for translation speed, and V_{rot}, for rotation speed) considering both reference commands and detected obstacles to be avoided.

Obstacles are detected using the laser sick range sensor that provides 720 data values for 180 degrees, where each datum gives the distance from the sensor to the nearest object detected. This is too much data to be processed, so we decided to reduce the 180 degrees of sensor detection to eight zones of 20 degrees each. Each zone contains 90 measures, but only the smallest one is used.

Figure 8.10 shows that each zone is in fact considered one object located at the smallest distance measured. d_1 is the distance of the obstacle in zone 1, and if no obstacle is present, this measurement will have the highest possible value.

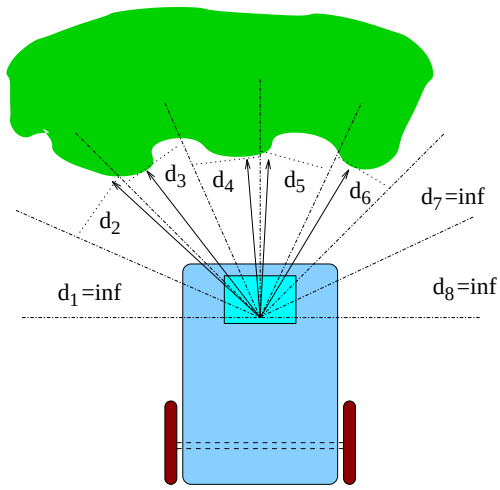


Figure 8.10: Zones situated in front of the robot with an obstacle and detected distances

The obstacle avoidance method proposed here consists of one Bayesian Program for each zone, and another Bayesian Program to combine the indications of all zones. The Bayesian program for each zone determines the robot commands to avoid the obstacle detected in this zone. The fusion program then integrates commands from all zones and reference values from the upper-level module.

The method employed here was developed initially for a car-like robot and was described by Koike et al. [2003b,a]. Later studies linked obstacle avoidance to navigation [Pradalier et al., 2003c, 2004, 2005].

Zone Bayesian Program

The distance measured in a zone is a sensor observation, indicating how close the detected object is in this zone. A distant object can be ignored, and then no special motor

command is recommended.

A close object demands deviation: depending on its position, some directions are forbidden; depending on its distance, the robot must move more slowly or even stop completely.

The Bayesian program for zone i uses three relevant variables: the sensor measure D_i, the rotation speed V_{rot} and the translation speed V_{trans}. Rotation and translation speed variables have the same type as M_{rot} and M_{trans} used in the Bayes filter. D_i variable values are the distance in centimetres, from zero to 20: if an object is beyond two metres , D_i is equal to 20.

The joint distribution is:

$$\begin{aligned} &P(V_{rot}\ V_{trans} D_i|\pi_i) \\ &\quad = P(D_i|\pi_i) \times P(V_{rot}|D_i\pi_i) \times P(V_{trans}|D_i\pi_i). \end{aligned} \tag{8.10}$$

No a priori information about where the objects can be located results in a uniform distribution for $P(D_i|\pi_i)$. The rotation speed depends on the object's distance, because a very close object requires a higher rotation speed. The curve for $P(V_{rot}|D_i\pi_i)$ for zone 3 is given in Figure 8.11. It can be seen that rotation speeds leading the robot towards the obstacle are forbidden. The translation speed is higher when the object is distant, and lower when it is close, as shown in Figure 8.11. This curve is employed for all zones.

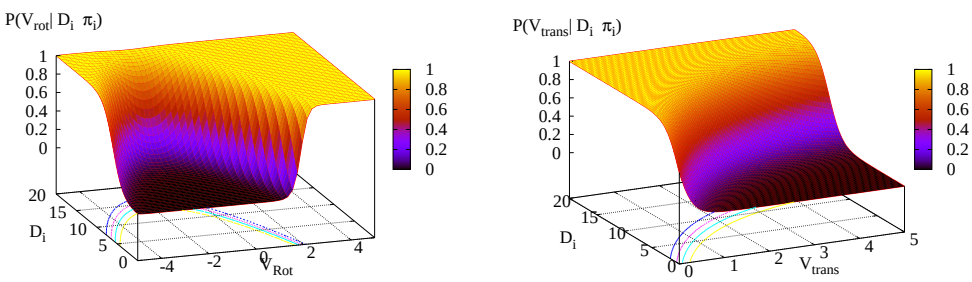


Figure 8.11: Rotation speed and translation speed terms in the zone bayesian program.

It is important to note that these curves do not indicate which direction to follow but rather impose constraints in relation to the position of an obstacle in the zone.

Utilisation of this program consists of the following question:

$$P(D_i|V_{rot}V_{trans}\pi) \propto P(V_{rot}|D_i\pi_i) \times P(V_{trans}|D_i\pi_i). \tag{8.11}$$

This question will be used by the fusion program.

Fusion Bayesian Program

In this program, motor commands from all zones are combined with reference values from upper-level modules. Relevant variables here are then: the set of sensor measures of the eight zones $D_1 D_2 \ldots D_8$; the rotation speed V_{rot}; the translation speed V_{trans}; M_{rot}, the rotation speed reference variable; and M_{trans}, the translation speed reference variable.

All commands proposed by the zones are combined with the reference commands by the joint distribution shown below:

$$\begin{aligned} & P(M_{rot}\ M_{trans}\ V_{rot}\ V_{trans}\ D_1 \ldots D_8 | \pi) \\ & = \left| \begin{array}{l} P(M_{rot} M_{trans} | \pi) \times P(D_1 \ldots D_8 | \pi) \\ \times P(V_{rot} | M_{rot} \pi) \times P(V_{trans} | M_{trans} \pi) \\ \prod_{i=1}^{8} P(D_i | V_{rot} V_{trans} \pi) \end{array} \right. \end{aligned} \tag{8.12}$$

There is no a priori knowledge about the reference values for rotation and translation speed in terms of the position and direction of obstacles, so the terms $P(M_{rot} M_{trans} | \pi)$ and $P(D_1 \ldots D_8 | \pi)$ are uniform distributions.

The next terms represent a fusion of the reference influence and the obstacle avoidance evidence from each zone. The rotation speed probability distribution term $P(V_{rot} | M_{rot} \pi)$ is a Gaussian; its mean value is the reference for rotation speed with constant standard deviation. The same principle is applied to the translation speed term $P(V_{trans} | M_{trans} \pi)$.

Terms $P(D_i | V_{rot} V_{trans} \pi)$ are actually questions asked of each zone joint distribution, as in equation 8.11.

Utilisation of the above joint distribution consists of asking the question:

$$P(V_{rot}\ V_{trans} | M_{rot}\ M_{trans}\ D_1 \ldots D_8\ \pi),$$

where M_{rot} and M_{trans} values are the result of the Bayes filter drawn from the motor commands global distribution, and $D_1 \ldots D_8$ are calculated knowing the measures from the laser range sensor.

8.3 Results

The specification was evaluated initially by the basic behaviours separately: chase, capture, escape, motionless, obey and wander. When problems were observed at this stage, changes were limited to the motor model; more precisely, the relation between the motor command variables and the state variables, as behaviour variables will have a fixed value.

Next, combinations of behaviours were evaluated; initially, basic behaviour with obstacle avoidance situations, and then behaviour coordination. This aimed to evaluate the behaviour model in each filter, followed by fusion of the answers to the behaviour questions and fusion of the motor question answers. Tuning involved changing values of probability, and eventually the tables, as described in Section 6.1.3.

The results shown here include only two basic behaviours (Sections 8.3.1 and 8.3.2) two situations showing coordination of basic behaviours, in Sections 8.3.3 and 8.3.4. In addition, obstacle avoidance is presented in the context of execution of a basic behaviour. The sequences presented below were performed in different rooms, with varied lighting conditions.

Section 8.3.5 analyses the results of this robot experiments, especially in relation to the applied programming methodology. Some shortcomings and drawbacks are presented and discussed.

8.3.1 Chase and Capture

Specified behaviour for the robot includes to chase and capture a prey. Once the prey is detected, the camera is centred on the prey image. Even while the robot is avoiding an obstacle, the camera remains fixed on the prey.

The robot's speed is limited for safety reasons. It accelerates only when the detected prey is distant. As the prey comes closer, the speed reduces. When the prey is detected to be very close to the robot (bottom middle snapshot in Figure 8.12), the robot stops and captures the prey with the laser pointer (last picture in Figure 8.12).

Results for chase behaviour executing in robot are shown in figure 8.12. Initially, top left snapshot, the robot wanders in environment and it detects a prey at left. Then, the robot goes towards the prey until it is close enough to capture it, activating the laser pointer.

8.3.2 Obey with Obstacle Avoidance

Obey behaviour consists in following the trainer: differently from chase behaviour, the robot keeps the distance and never captures the trainer with laser pointer.

Figure 8.13 shows how obey behaviour is executed by the robot. The robot detects the trainer and follows it, as shown in the first two pictures. An obstacle is between the robot and the trainer, and the robot bypasses it, always keeping the camera fixed on the trainer, as shown in the last three pictures.

8.3.3 Reactivity to Predator

In the specified behaviour, an important requirement is to be reactive to predator presence, and react accordingly: stay motionless, if detected predator is faraway; escape if detected predator is near; or follow the trainer, if it is present to assure the robot safety.

Figure 8.14 displays a situation where the behaviour implemented in robot reacts to predator presence. Initially, chases a prey (first two pictures). Then the predator arrives, and it is very close to the robot, which makes a left turn and escapes from the predator. While following the farthest direction, the robot is trapped in a corner, as if it was hiding from the predator.

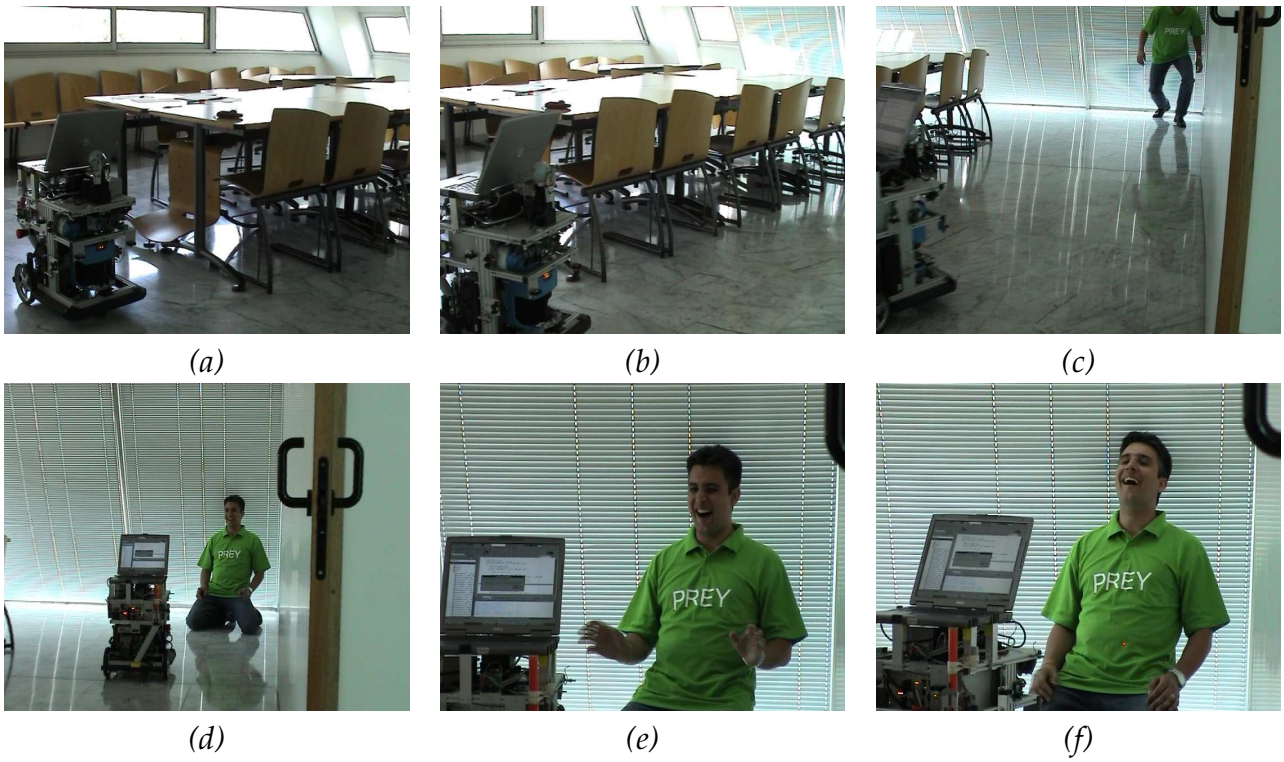


Figure 8.12: Chase behaviour. From top left to bottom right: (a) the robot wanders in environment, avoiding obstacles; (b) it detects a prey at its left side; (c) the robot advances towards prey; (d) the robot approaches nearest possible; (e) the prey is within reach; (f) the prey is captured with laser pointer

8.3.4 Behaviour Coordination

It is desired to illustrate results of robot implementation when basic behaviours are combined, especially conflicting ones. The situation shown in figure 8.15 begins with the robot wandering, when a predator comes into the robot's field of view (first row of snapshots). As the predator is distant, the robot decides to stay motionless.

Next, the trainer arrives, and the robot follows him even though the trainer moves and stays close to the predator. Even while following the trainer, the robot avoids going directly towards the predator.

8.3.5 Result Analysis

Comparing Robot Resulting Behaviour with Specified Behaviour

Robot resulting behaviour was evaluated regarding execution of basic behaviours and their coordination, and it is considered to be compatible with specification description in section 8.1.1. Desired characteristics, as reactivity to predator appearance and persistence in chasing and obeying, were observed in several situations.

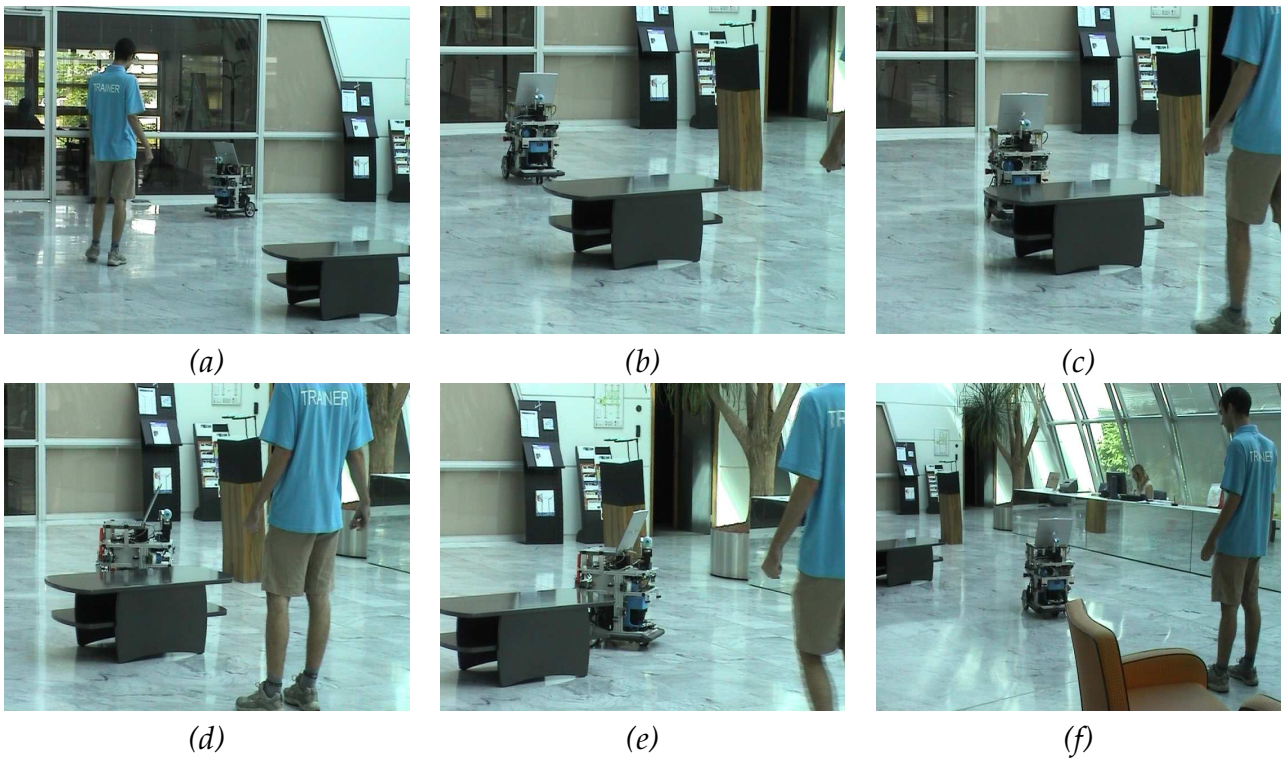


Figure 8.13: Obey behaviour, with obstacle avoidance. From top left to bottom right: (a) the robot detects trainer; (b) the robot follows the trainer movements; (c) an obstacle comes between the robot and the trainer; (d), (e) the robot changes direction to avoid obstacle, but camera is always fixed on the trainer; (f) the robot advances towards the trainer.

One of the disadvantages of the implementation shown here is the limited speed of the robot. The robot behaviour involves interaction with humans, and for safety reasons, it must have a limited speed. However, the present speed is well below the desired and allowed value: motor decisions are taken each 250 milliseconds, but 100 milliseconds is the desired maximum duration between motor command decisions. This value is considered reasonable for a robot evolving in an environment with persons.

Exact probabilistic calculations demand no more than 90 milliseconds, so no approximations were found necessary. The main restriction is in relation to vision capture and processing.

The camera is mounted on a pan–tilt head, as shown in Figure 8.4. A capture-processing cycle involves capturing an image frame, processing it, and moving the pan–tilt head to centralise the biggest object detected, demanding 200 miliseconds. After moving the head, it is necessary to wait at least 50 milliseconds so that the next image capture is stable enough to be processed. In summary, between two captures of image frames, 250 milliseconds.

In the present situation, the selective perceptive method described in Chapter 7

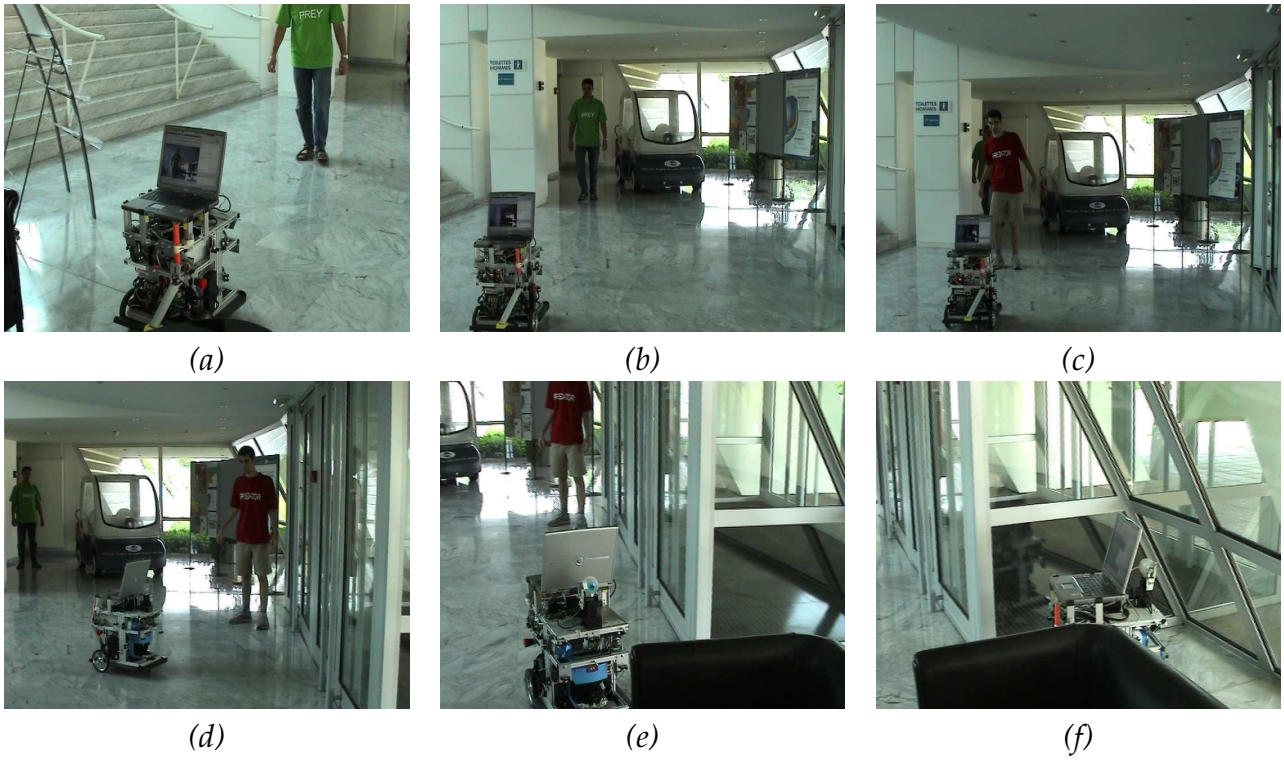


Figure 8.14: Chase behaviour, with reactivity to predator appearance. From top left to bottom right: (a) the robot detects a prey; (b) the robot advances towards detected prey; (c) the robot detects a predator; (d), (e) escaping from detected predator; (f) hiding from detected predator (actually, blocked in a corner).

would not bring any advantage: processing one object instead of three would decrease the time between two captures to 230 milliseconds, still far from the desired 100 milliseconds.

Programming Methodology

The theoretical structure of the filters presented in Chapter 6 seems intricate, but implementation in a real robot is not particularly difficult.

One of the basic assumptions of the Bayesian robot programming approach is that the *description* and *utilisation* phases are independent. When writing terms in the joint distribution decomposition, only previous knowledge about context and the situation being modelled is employed: it is important not to preview utilisation. The same program description can answer several questions, regarding any of its relevant variables.

In spite of number of involved variables and joint distribution's size (see equation 8.1), the programming task is simplified by the independence of the elementary filters. Assumptions necessary to have dissociate elementary filters are hard to fulfil, but once these conditions are met, the consequent independence for programming each filter is

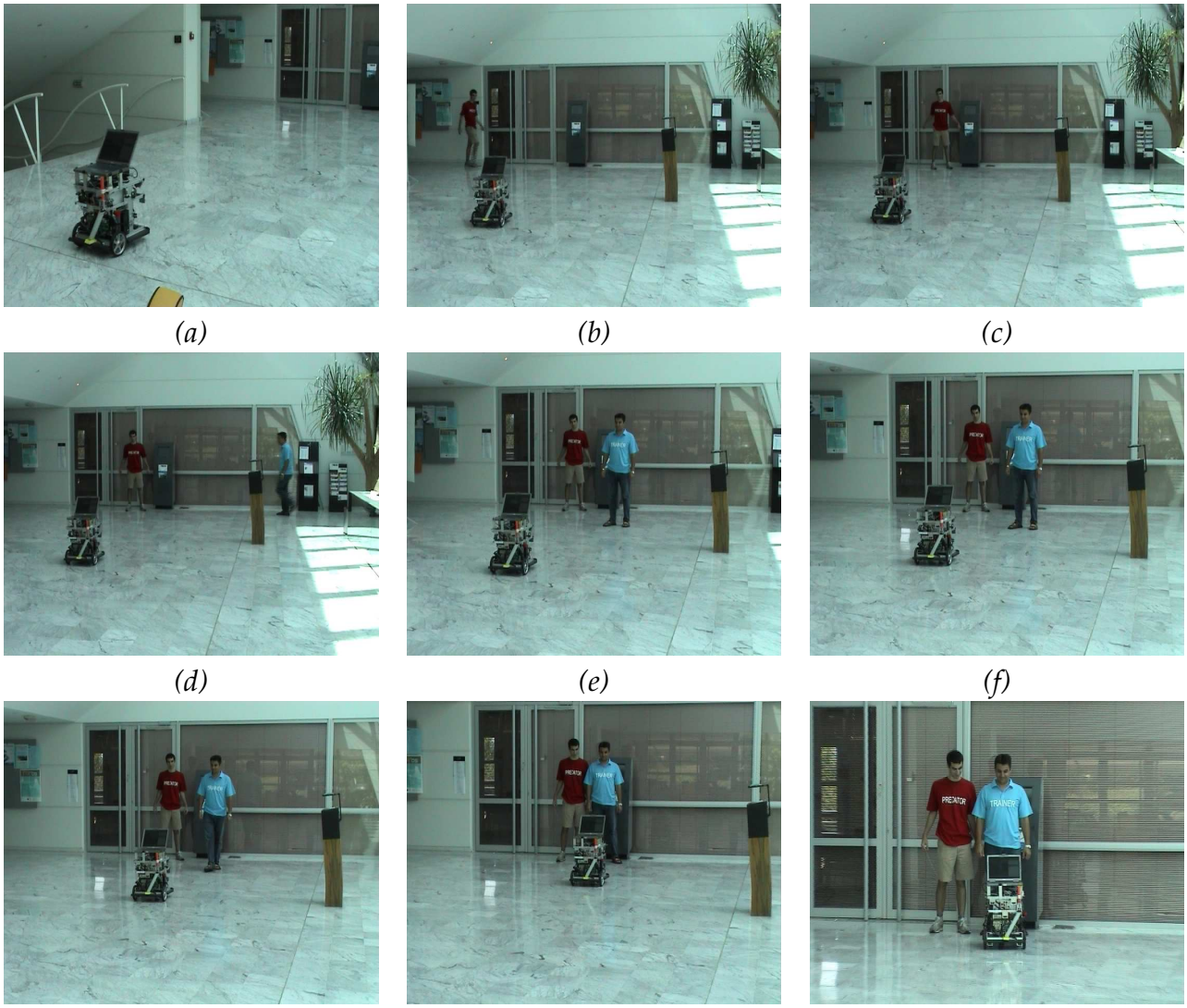


Figure 8.15: Behaviour coordination. From top left to bottom right: (a), (b) the robot wanders; (c) it detects a far predator; (c), (d) it keeps motionless; (e) approaching trainer detected; (f),(g),(h) the robot follows detected trainer; (i) even if a predator is very near, stays with the trainer.

very useful. Each term in the joint distribution is then programmed employing the knowledge provided by the dependent variables.

Regarding terms involving common variables (behaviour and motor commands, for example), programming can become a delicate task. Because the robot's resulting behaviour comes from global distributions over common variables, it appears that it might be necessary to balance the probability values within each elementary filter so that the global distributions correspond to the specified behaviour.

In fact, the elementary filters independence is also essential when dealing with common variables, mainly because coherence-based fusion of common variables provides a satisfying combination of elementary proposals. Terms involving common variables in each elementary filter must therefore be filled considering only local information, as state variables. Situations where local information is enough to recommend or forbid a specific value for a common variable are easily programmed: recommended values have high probability values, and forbidden ones have very low probability values.

In other situations, local information is not sufficient. It is important neither to prohibit a behaviour (as another filter could indicate it as being interesting), nor to recommend it (as it is not possible to know if it is really indicated). Medium probability values are the most appropriate in these cases.

The method proposed in Chapters 6 and 7 to fill the behaviour and attention selection models is coherent with the reasons above, and it is effective in reducing complexity.

Behaviour implemented in robot used a global joint distribution with thirty six terms to be programmed. Considering that the reasoning is similar for some terms, only fifteen different models exists, employing programming guidance proposed here.

Practical Utilisation of Proposed Framework

The proposed framework was applied in a real robot application, using a programming methodology to define a priori all free parameters, and the resulting behaviour is acceptable in relation to desired behaviour specification. This application makes also possible to examine the filter working details in a realistic situation.

Application implementation consists in codifying the expressions solving the questions and filling the parametrical forms and tables as exposed in Section 8.2.1. Surely, vision processing and communication with robot sensors and actuators must also be implemented, but regarding behaviour control system, no aditional skill is necessary.

A summary of all strong and weak points in the application of proposed framework in the experiment described in this chapter is presented in table 8.8. They are associated to some of the main points discuted in Chapters 5 to 7, and how they were applied in the experiment.

Key Points	**Strengths**	**Shortcomings**
Disjoint subsets of state and observation variables	Variables are associated to subjects in the environment related to specified behaviour	Interactions between subjects are difficult to taken into account.
Employing Fusion with coherence	Easily written from an inverse model	Requires definition in a table of probabilities.
Repertoire of simple and independent motor patterns	Motor patterns are directly associated to action towards environment subjects used in state variables subsets and easy to program.	Repeated patterns if the same action happens towards different subjects.
Behaviour variable switches between motor patterns	Programming of motor model becomes very simple	Only one behaviour is active at a time, and compromise solutions are obtained in behaviour level, not in motor model.
Behaviour choice is the result of the fusion of Elementary filter proposals	Proscriptive programming simplify writing local models.	Requires discipline to employ only available knowledge, when programming behaviour models.
Addition of attention model	Three subjects are detected by vision and, with attention focusing, processing time can be reduced to one third	Camera centralisation in the subject and image stabilisation is more time consuming than processing: focusing attention is not sufficient for reducing detection time
Proscriptive Programming	Widely employed in behaviour and motor models	Requires combination with prescriptive programming, otherwise it can result in uniform distributions

Table 8.8: Summary of strengths and shortcomings of proposed framework key points in the experiment described in this chapter.

Chapter 9

Conclusion and Perspectives

9.1 Contributions and Main Results

In this thesis, we propose a succession of assumptions and simplifications related to controlling autonomous sensory-motor systems. Our main contribution is a framework for robot programming based on strict mathematical definitions and featuring Bayesian-based action selection and selective perception.

The framework joint distribution is shown in Figure 9.1, and each model in this distribution is enclosed by a coloured rectangle: green for the dynamic model; blue for the sensor model; red for the behaviour model; orange for the attention model; brown for the motor commands; and yellow for the initial conditions. Magenta circles indicate

$$
\begin{aligned}
&P(M^{0:t} S^{0:t} Z^{0:t} B^{0:t} C^{0:t} \lambda^{0:t} \beta^{0:t} \alpha^{0:t} | \pi) \\
&= \left| \prod_{j=1}^{t} \left[\begin{array}{l}
\boxed{\prod_{i=1}^{N_i} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i)} \\
\boxed{\prod_{i=1}^{N_i} P(Z^j | S_i^j C^j \pi_i)} \\
\boxed{\times P(B^j | \pi) \prod_{i=1}^{N_i} P(\beta^j | B^j S_i^j B^{j-1} \pi_i)} \\
\boxed{\times P(C^j | \pi) \prod_{i=1}^{N_i} P(\alpha^j | C^j S_i^j B^j \pi_i)} \\
\boxed{\times P(M^j | \pi) \prod_{i=1}^{N_i} P(\lambda^j | M^j B^j S_i^j M^{j-1} \pi_i)}
\end{array} \right] \right. \\
&\quad \boxed{\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi)}
\end{aligned}
$$

Figure 9.1: Joint distribution of the proposed framework. The green square indicates the dynamic model; the blue square indicates the sensor model; the red square indicates the behaviour model; the orange square indicates the attention model; the brown square indicates the motor commands model; and the yellow square indicates initial conditions.

elementary filter products.

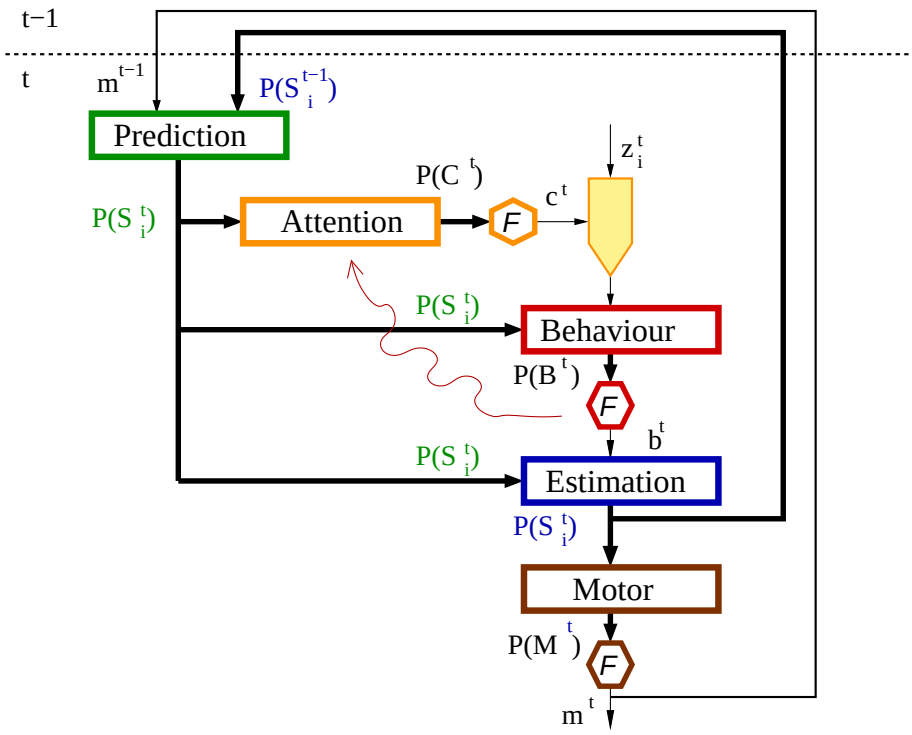


Figure 9.2: Utilisation of Bayes filter with behaviour coordination and attention selection.

Utilisation of the joint distribution is defined as shown in Figure 9.2. This shows that asking questions of the above joint distribution is equivalent to asking questions of elementary filters with the joint distribution:

$$
\begin{aligned}
& P(M^{0:t} S_i^{0:t} Z_i^{0:t} B^{0:t} C^{0:t} \lambda_i^{0:t} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i) \\
& = \left| \begin{array}{l} \Pi_{j=1}^{t} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j C^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\ \times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i). \end{array} \right.
\end{aligned}
\tag{9.1}
$$

In Figure 9.2, the dotted line indicates the limit between two time instants. The first question asked of the joint distribution in equation 9.1 is the prediction question, indicated by the green box. In the joint distribution, the dynamic model is also indicated by a green rectangle: actually, the dynamic model is employed in the prediction question expression to update the probability distributions over the state variables.

Then, the attention question, inside the orange box, is asked of this same joint distribution. Fusion of elementary filter answers to the attention question is indicated by the orange hexagon: from the fusion result, values are drawn for attention variables, which influence sensor data processing, and consequently the observations available. The attention model in the joint distribution is also enclosed by an orange rectangle, as it is directly connected to the attention question expression.

The next question asked of the elementary filter joint distribution is the behaviour question, indicated by the red box, followed by fusion of elementary filter answers, shown by the red hexagon. From the resulting probability distribution, values for behaviour variables are drawn. The behaviour model in the joint distribution is shown by a red square, because of its importance in the expression answering the behaviour question.

The estimation question (blue rectangle) is then asked of the joint distribution in equation 9.1, followed by the motor question, shown by the brown square. Fusion of elementary filter answers to the motor question is then executed, as indicated by the brown hexagon, and from the fusion result, values for the motor command variables are drawn. These drawn values, as well as the result from the estimation question, are used in the next time instant to calculate the prediction question.

The blue box in the joint distribution encloses the sensor model, as it is related to the estimation question answer even in the simpler Bayes filter. The motor commands model is indicated by the brown colour in the joint distribution because of its closer relation to the motor command question.

Secondary contributions of this work concern the solutions employed in this framework, the analysis and discussion included in this document, and the programming method applied in the simulated and robot experiments presented, as detailed in the following.

Addition of Motor Command in the Bayes Filter Recursive Loop of Calculations

As shown in Chapter 4, the motor command model is introduced in the joint distribution of the Bayes filter, and inference of motor commands is added in the recursive loop of state estimation. This is a simpler alternative to POMDP action inference, as there is no optimisation and no time horizon. It allows decisions to be taken very quickly, to cope with fast-changing environment conditions.

Coherence-based fusion was employed to ensure inference of motor commands in the case-disjoint subsets of state and observation variables, to allow utilisation of elementary filters.

Action Selection Mechanism is Included in the Bayes Filter

A mechanism for selection of behaviour is added to the Bayes filter in Chapter 6, profiting from the recursive calculation loop. Action selection is programmed in the behaviour selection model included in the joint distribution, and selection of adequate be-

haviour is carried out inside the recursive loop of the state estimation by means of the behaviour questions.

The behaviour model benefits from the simplified programming of coherence-based fusion if the conditions for elementary filter utilisation are met.

Bayesian Selective Perception Associated with Action Selection

A selective perception mechanism is added to the Bayes filter in Chapter 7, and it is proposed that it is strongly linked to the action selection mechanism.

The attention selection model is added to the filter joint distribution, and the dependence on the behaviour variable defines the relation between attention and behaviour. The attention question is included in the filter loop of the state estimation, and the expression answering this question provides predictions for behaviour selection . The attention selection model affects how sensor data is processed, which observations are available, and which sensor models are applied.

Analysis of the Time and Space Complexities of the Bayes Filter Description and Utilisation

The influence of the description and utilisation concepts of the Bayes filter on time and space complexity is examined. This analysis is executed in all chapters for all proposed filters.

Proposals are compared under similar conditions, and no previous assumptions about number of variables or application-dependent relations are considered. Additionally, some assumptions regarding robotics applications are used to estimate the real costs of adding action and attention selection mechanisms to the robot control system.

Discussion of Cognitive Aspects of the Bayes Filter Description and Utilisation

The proposed Bayes filter is analysed in relation to cognitive features linked to autonomous sensory-motor systems behaviour, for all proposed filters.

Programming Method

Because of the complexity of the proposed filter, techniques for programming a specified behaviour are of the utmost importance. Some convenient rules of thumb help to program terms in the joint distribution.

First, the Bayesian program description expresses preliminary knowledge about the system: robot, behaviour and environment. After definition of the Bayesian program description, utilisation happens independently. Decomposition and parametric forms must be expressions of knowledge about the problem and cannot depend on utilisation questions.

Once decomposition is chosen, each term in the joint distribution is considered separately, and only variables concerned in each term must be taken into account.

This rule is especially useful when elementary filters are employed. Each elementary filter has its own joint distribution: the terms in this joint distribution are defined using only the knowledge available in the elementary filter, even if they are related to common variables such as behaviour, attention or motor commands.

Proscriptive programming must be used wherever possible, and it is helpful to identify distinctly proscriptive behaviours. When fusion (either sensor fusion or coherence-based fusion) is employed, it is important to define clearly prohibited situations (by assigning them very low probability values) and to give appropriate probability values to other situations: unconcerned, not relevant or unwise.

As discussed for the illustrative examples in Chapters 6 and 7, levels of probability value can be associated with situations discerned for selection of behaviours or focuses of attention. This can substantially reduce the number of parameters requiring tuning.

9.2 Perspectives

Mapping

Most studies of autonomous robotics consider mapping and planning essential for navigation [Thrun, 2002b]. In relation to the framework proposed in this thesis, utilisation of maps is possible, although it can be particularly difficult to respect the constraints suggested in Chapter 5.

The illustrative examples employed in several chapters did not mention any form of maps, but a very simple map was implemented and tested together with the examples from Chapters 5 to 7.

As the simulated environment is very simple, the map employed is similar to the one proposed by Diard [2003, chap. 6]: locations are corner, wall and free space. The map was modelled as an additional elementary filter, with no behaviour or attention model. The motor command model proposes motor commands according to the present position in the map.

Extension of this simple example to a large-scale map leads to interesting problems. The large number of necessary state variables required for maps, which are usually highly connected models, is almost impossible to break into disjoint subsets. Can coherence-based fusion help in this situation, as with the attention, behaviour and motor variables presented in this work? Are Bayesian maps and their operators still valid in a system where the map representation is one of the elementary filters? Is planning feasible in this context?

Strongly Connected Models

Boyen [2002] proposed a compact representation (very similar to our disjoint subsets

proposal) to replace a full model with a larger number of dependencies. It was shown that it is possible to evaluate the error between the compact representation utilisation and the full system utilisation, and then verify if the representation can be used.

It would be interesting to verify the utilisation of the proposed method for finding compact representations in systems where it is not possible to find disjoint subsets of state variables, and to verify the error between the compact and full representations under our framework using the analysis techniques proposed.

Comparison with Other Action Selection Mechanisms

Tyrrell [1993] proposed a simulator to compare action selection mechanisms. The specified behaviour was more complex that those discussed here, and it is an interesting environment to compare our framework's efficacy.

Sensor signals can be better controlled in simulation, and it can be helpful to verify efficacy with and without selective perception associated with behaviour coordination.

We have emphasised cognition aspects present in the resulting behaviour of our proposed framework. No analysis was developed towards physiological models of action selection, as proposed by Girard et al. [2003, 2002]. Particularly interesting is the utilisation of stabilisation analysis techniques, such as contraction theory, of the basal ganglia model of action selection [Girard et al., 2005].

Active Perception

The selective perception included in this model does not consider that motor variables may be necessary to control perception data: to move the head of the robot to search for an object, for example.

Several studies include active perception in maps built in Kalman or particle filters, and the combination of active and selective perception is an interesting cognitive aspect to be investigated.

Temporal Loops and Shortcuts

Temporal relationships between the filters are necessary for certain situations: defence-related filters, for example, must provide fast answers, and a kind of temporal shortcut for defence calculations can be conceived.

There are situations where some states are not concerned with presence behaviour. To reduce calculation time, only the filters concerned may be active: unconcerned filters should not execute as many calculation loops as concerned filters. This is very similar to the concept of navigation through macro actions as proposed in Theocharous and Kaelbling [2003].

9.2.1 Hierarchical Bayesian Program

Hierarchy in behaviour is important for more complex specified behaviours [Prescott et al., 1999; Bryson, 2000]. Figure 9.3 shows a diagram of a defence-related behaviour hierarchy, defined using the behaviour systems approach [Timberlake and Silva, 1995]. This approach was applied to model and program a mobile robot by Arkin et al. [2001], and a nice discussion about its basic assumptions can be found in Tyrrell [1993, chapter 8].

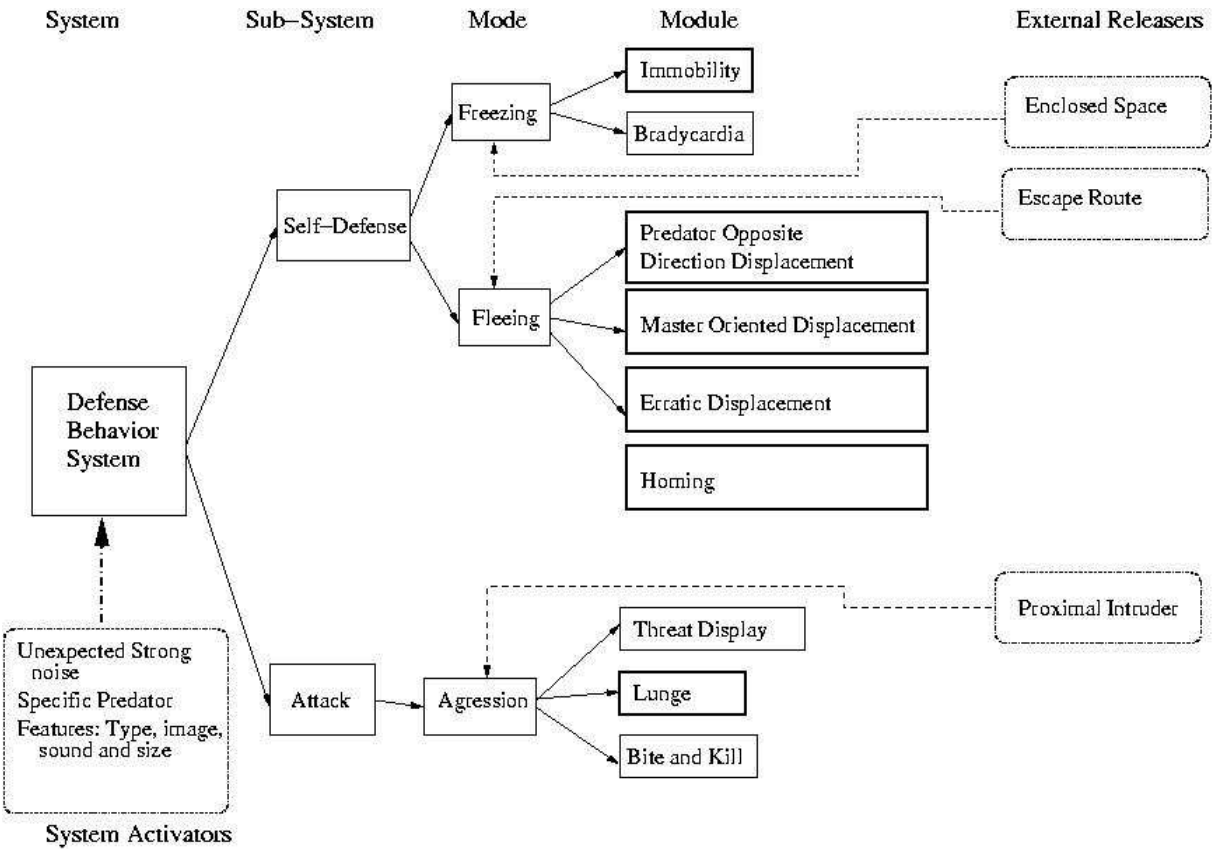


Figure 9.3: Hierarchical model for defence behaviour applying the behaviour systems approach (see Timberlake and Silva [1995] for structure description).

Extending the filter proposed in Chapter 7 to implement hierarchical specified behaviour can be done in several different ways. Below, work in progress towards a hierarchical structure of behaviour and attention selection is presented.

Description

Relevant Variables

Variables are the same as described in Chapter 7. However, behaviour and attention variables receive an additional index: ${}_{[k]}C^t$ means the set of attention variables at level k at time instant t, while ${}_{[k]}B^t$ stands for the set of behaviour variables at level k at time instant t.

Observations also depend on the hierarchical level, and observation variables also receive an additional index: ${}_{[k]}Z_i^t$ stands for the set of observation variables of level k in elementary filter i at time instant t.

Joint Distribution

Each elementary filter has a joint distribution like:

$$
\begin{aligned}
&P(S_i^{0:t} \wedge {}_{[1:L]}C^{0:t} \wedge {}_{[1:L]}Z_i^{0:t} \wedge {}_{[1:L]}B^{0:t} \wedge M^{0:t} \wedge \lambda_i^{0:t} \wedge {}_{[1:L]}\beta_i^{0:t} \wedge {}_{[1:L]}\alpha_i^{0:t} \,|\pi_i) \\
&= \left| \prod_{j=1}^{t} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \right. \\
&\left. \times P(S_i^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i) \right.
\end{aligned}
$$

Inside the time product, the first line shows the dynamic model, followed by four lines that contain the behaviour model, the attention model, and the sensor model for the first (also called the top) level in the hierarchy . A product of models for all other levels is shown below. Here, the dependence on the behaviour model with the behaviour variables in the previous level is explicit in:

$$P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}B^{t-1}\pi_i).$$

This dependence is the only reason why models in level one are separated from models of other levels: level one has no previous level to depend on.

The next line inside the time product shows the motor model. Only behaviour variables from the last level appear in this model:

$$P(\lambda_i^t|M^t \wedge M^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i).$$

Behaviour, attention and motor models are defined by applying coherence- based fusion because they are related to variables common for all filters in a specific level.

Bayesian network representation of elementary filter joint distribution in presented in appendix E, Figure E.1.

Utilisation

The hierarchy in behaviour and perception is visible in the utilisation of each elementary filter. The definition of one hierarchical level for each elementary filter is presented in Figure 9.4.

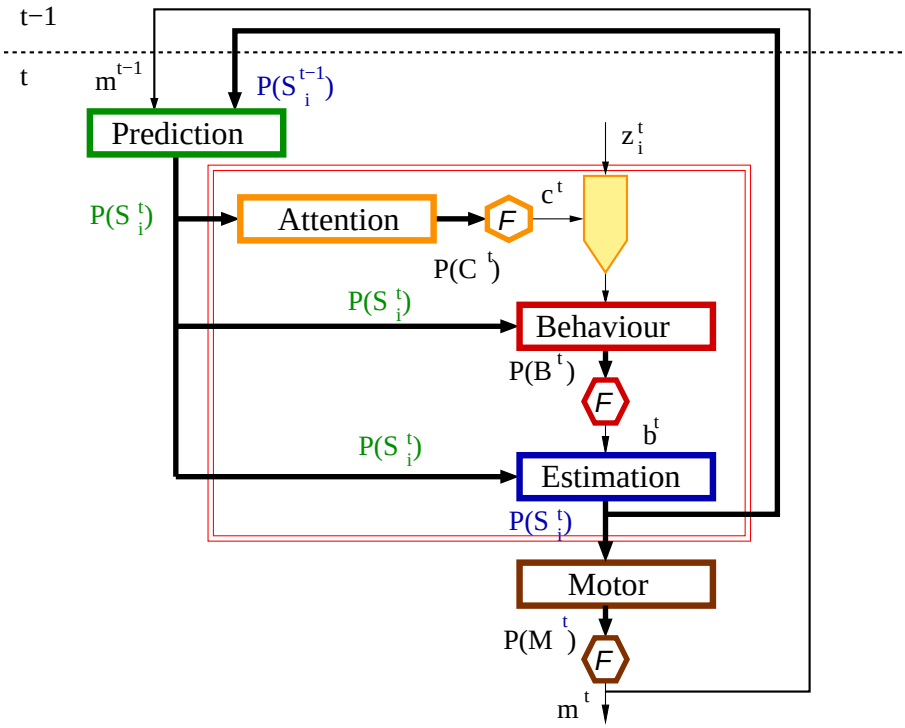


Figure 9.4: Definition of one level in the hierarchy.

We can see that each level contains one attention question, one behaviour question and one estimation question. This means that these three questions are asked at all hierarchical levels: attention decides which relevant new information about the environment should be observed; behaviour decides which of the possible actions is selected, and estimation is necessary to maintain coherence between state predictions and decisions on attention and behaviour.

Estimation is then applied by the next hierarchical level (or by the motor model) as a priori information about state variables. The drawn values for the behaviour variables are also necessary at the next level. Figure 9.5 illustrates a system with three hierarchical levels in one elementary filter.

The prediction question in an elementary filter is asked once for each time step, followed by a block of questions regarding the first level in the hierarchy.

The attention question in each elementary filter is calculated, and the answers are combined. From the resulting distribution, a value is selected for variable ${}_{[1]}C^t$. This value is used to identify the observations available in observation set ${}_{[1]}Z_i^t$ when asking the behaviour question of each elementary filter. Answers to the behaviour questions

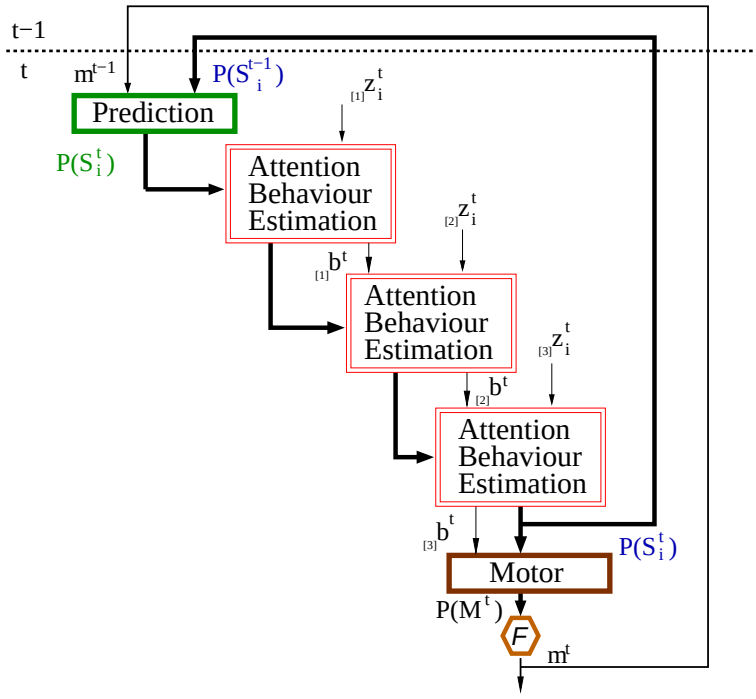


Figure 9.5: Utilisation of hierarchy levels in an elementary filter.

are combined and a value for variable $_{[1]}B^t$ is decided. Using the selected values for attention and behaviour at this level, the estimation question for this level is asked.

The next level of calculations then takes place: first the attention question, then behaviour, and last estimation. Levels succeed until the last one: only then is the motor question asked of each elementary filter, answers are combined and values for motor command variables are drawn.

For details regarding expressions answering these questions see Appendix E.

Main Assumptions

- Complex behaviours require a hierarchy to be described.
- Each level is influenced by decisions in the previous level.
- Each level in behaviour selection has specific observation requirements.
- The choice of motor commands depends only on the lowest level of behaviour selection.

9.3 Conclusion

This document has presented a framework for programming a robot control system to execute a previously specified behaviour. The proposed framework consists in an im-

proved Bayes filter, including selective perception, hierarchical behaviour selection and safe motor control and it is defined within a precise and strict mathematical framework.

A succession of cumulative hypotheses and simplifications have been presented in five stages. The validity of these hypotheses can be discussed and challenged. However, the corresponding inferences and resulting decisions are derived mathematically.

Each stage description has been followed by its analysis according to memory requirement, processing complexity, and difficulty of modelling. Further discussions regarding robot programming and cognitive modelling points of view have also been presented.

An implementation on a mobile robot has been described, whose results demonstrate that the proposed framework is adequate for practical purposes.

Bibliography

Y. Aloimonos, editor. *Active Perception*. Lawrence Erlbaum Associates, Inc., 1993. ISBN 0805812903.

R. Arkin, M. Fujita, T. Takagi, and R. Hasegawa. Ethological modeling and architecture for an entertainment robot. In *Proceedings of ICRA 2001*, 2001.

S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, Feb. 2002.

J.-C. Baccon, L. Hafemeister, and P. Gaussier. A context and task dependent visual attention system to control a mobile robot. In *Proceedings of the 2002 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, pages 238–243, Lausanne, Switzerland, October 2002. EPFL.

D. Bellot, R. Siegwart, P. Bessière, A. Tapus, C. Coué, and J. Diard. Bayesian modeling and reasoning for real world robotics: Basics and examples. In F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, editors, *Embodied Artificial Intelligence, Int. Seminar, Dagstuhl (DE), July 7-11, 2003*, volume 3139 of *Lecture Notes in Computer Science*, pages 186–201. Springer-Verlag, 2004.

N. Bernstein. *The coordination and regulation of movements*. Pergamon, London, 1967.

A. Berthoz. *The brain's sense of movement*. Harvard University Press, 2000.

A. Berthoz. *La Déision*. Editions Odile Jacob, 2002.

P. Bessiere and B. I. R. Team. Survey: Probabilistic methodology and techniques for artefact conception and development. Technical Report RR-4730, INRIA, 2003.

P. Bessière, E. Dedieu, O. Lebeltel, E. Mazer, and K. Mekhnacha. Interprétation ou description (I): Proposition pour une théorie probabiliste des systèmes cognitifs sensori-moteurs. *Intellectica*, 26–27(1–2):257–311, 1999a.

P. Bessière, E. Dedieu, O. Lebeltel, E. Mazer, and K. Mekhnacha. Interprétation ou description (II): Fondements mathématiques de l'approche F+D. *Intellectica*, 26–27(1–2): 313–336, 1999b.

J. Binder, D. Koller, S. Russell, and K. Kanazawa. Adaptive probabilistic networks with hidden variables. *Machine Learning*, 29:213–244, 1997.

X. Boyen. *Inference and Learning in Complex Stochastic Processes*. PhD thesis, Stanford University, December 2002.

D. Bryant. A spatial representation system in humans. target paper in Psycoloquy, psycoloquy.92.3.16.space.1.bryant, psyc@pucc.bitnet, 1992.

J. Bryson. *Intelligence by Design: Principles of Modularity and Coordination for Engineering Complex Adaptive Agents*. Ph.D. Thesis, Massachusetts Institute of Technology, Massachusetts, Sept. 2001.

J. J. Bryson. The study of sequential and hierarchical organisation of behaviour via artificial mechanisms of action selection, 2000.

A. R. Cassandra, L. P. Kaelbling, and M. L. Littman. Acting optimally in partially observable stochastic domains. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, Seattle,USA, 1994.

C. Coué. *Modèle bayésien pour l'analyse multimodale d'environnements dynamiques et encombrés : application à l'assistance à la consuite automobile en milieu urbain.* Thèse de doctorat, Inst. Nat. Polytechnique de Grenoble, Dec. 2003.

F. Crabbe. Optimal and non-optimal compromise strategies in action selection. In *Proceedings of the Eighth International Conference on Simulation of Adaptive Behavior*, Los Angeles, USA, July 2004.

F. Crabbe. On compromise strategies for action selection with proscriptive goals. In *Proceedings of Modeling Natural Action Selection*, Eddinburgh, Scotland, may 2005.

A. Davison. *Mobile Robot Navigation using Active Vision*. PhD thesis, Department of Engineering Science, University of Oxford, 1998.

T. Dean, L. P. Kaelbling, J. Kirman, and A. Nicholson. Planning under time constraints in stochastic domains. *Artificial Intelligence*, 76(1-2):35–74, July 1995.

F. Dellaert, S. Seitz, C. Thorpe, and S. Thrun. EM, MCMC, and chain flipping for structure from motion with unknown correspondence. *Machine Learning*, 50(1-2), 2003.

J. Diard. *La carte bayésienne : un modèle probabiliste hiérarchique pour la navigation en robotique mobile*. Thèse de doctorat, Inst. Nat. Polytechnique de Grenoble, Jan. 2003.

J. Diard, P. Bessière, and E. Mazer. A survey of probabilistic models, using the bayesian programming methodology as a unifying framework. In *Proc. of the Int. Conf. on Computational Intelligence, Robotics and Autonomous Systems*, Singapore (SG), Dec. 2003.

J. Diard and O. Lebeltel. Bayesian programming and hierarchical learning in robotics. In F. R. Meyer, Berthoz and Wilson, editors, *SAB2000 Proceedings Supplement Book*, Honolulu, 2000. Publication of the International Society for Adaptive Behavior.

D. Fox, W. Burgard, and S. Thrun. Active markov localization for mobile robots. *Robotics and Autonomous Systems*, 25(3-4):195–207, 1998a.

D. Fox, W. Burgard, S. Thrun, and A. Cremers. A hybrid collision avoidance method for mobile robots. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, volume 2, pages 1238–43, Leuven, Belgium, May 1998b.

B. Girard, V. Cuzin, A. Guillot, K. Gurney, and T. Prescott. A basal ganglia inspired model of action selection evaluated in a robotic survival task. *Journal of Integrative Neuroscience*, 2(2):179–200, 2003.

B. Girard, V. Cuzin, A. Guillot, K. N. Gurney, and T. J. Prescott. Comparing a brain-inspired robot action selection mechanism with winner-takes-all. In B. Hallam, D. Floreano, J. Hallam, G. Hayes, and J.-A. Meyer, editors, *From Animals to Animats 7. Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*. The MIT Press, 2002.

B. Girard, N. Tabareau, J.-J. Slotine, and A. Berthoz. Contracting model of the basal ganglia. In J. Bryson, T. Prescott, and A. Seth, editors, *Modelling Natural Action Selection: Proceedings of an International Workshop*, pages 69–76, 2005.

C. Guestrin. *Planning Under Uncertainty in Complex Structured Environments*. Ph.d. dissertation, Computer Science Department, Stanford University, August 2003.

M. Humphrys. *Action Selection Methods using Reinforcement Learning*. PhD thesis, University of Cambridge, Computer Laboratory, Cambridge, England, june 1997.

L. Itti. *Handbook of Brain Theory and Neural Networks*, chapter Visual Attention, pages 1196–1201. MIT Press, second edition edition, January 2003a.

L. Itti. Modeling primate visual attention. In J. Feng, editor, *Computational Neuroscience: A Comprehensive Approach*, pages 635–655. CRC Press, Boca Raton, 2003b.

E. Jaynes. *Probability theory - The logic of science*. Cambridge University Press, 2003.

L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 1998.

S. Koenig and R. Simmons. Unsupervised learning of probabilistic models for robot navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, Minneapolis (USA), 1996.

C. Koike, C. Pradalier, P. Bessière, and E. Mazer. Obstacle avoidance and proscriptive bayesian programming. In *Proc. of the Workshop on Reasoning with Uncertainty in Robotics*, Acapulco (MX), July 2003a.

C. Koike, C. Pradalier, P. Bessière, and E. Mazer. Proscriptive bayesian programming application for collision avoidance. In *Proc. of the IEEE-RSJ Int. Conf. on Intelligent Robots and Systems*, Las vegas, NV (US), Oct. 2003b.

T. Lane and L. P. Kaelbling. Toward hierarchical decomposition for planning in uncertain environments. In *Proceedings of the 2001 IJCAI Workshop on Planning under Uncertainty and Incomplete Information*, pages 1–7, 2001.

R. Le Hy, A. Arrigoni, P. Bessière, and O. Lebeltel. Teaching bayesian behaviours to video game characters. *Robotics and Autonomous Systems*, 47:177–185, 2004.

O. Lebeltel. *Programmation bayésienne des robots*. Thèse de doctorat, Inst. Nat. Polytechnique de Grenoble, Grenoble (FR), Oct. 1999.

P. Maes. How to do the right thing. *Connection Science Journal, Special Issue on Hybrid Systems*, 1, 1990.

N. Mansard. Hiérachie de comportements: application au retour à la base. Mémoire de Diplôme d'Etudes Approfondies, Inst. Nat. Polytechnique de Grenoble, Grenoble (FR), June 2003.

N. Mansard, O. Aycard, and C. Koike. Hierarchy of behaviours application to the homing problem in indoor environment. In *IEEE ROBIO 2005*, Hong Kong and Macau, China, July 2005.

A. K. McCallum. *Reinforcement Learning with Selective Perception and Hidden State*. Phd. thesis, University of Rochester, Rochester, USA, December 1995.

K. Mekhnacha, E. Mazer, and P. Bessière. The design and implementation of a bayesian CAD modeler for robotic applications. *Advanced Robotics*, 15(1):45–70, 2001.

L. Mihaylova, T. Lefebvre, H. Bruyninckx, K. Gadeyne, and J. D. Schutter. Active sensing for robotics - a survey. In *Proceedings of the Fifth International Conference on Numerical Methods and Applications*, 2002.

K. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California, Berkeley, California, July 2002a.

K. Murphy. Learning markov processes. In L. N. et al, editor, *The Encyclopedia of Cognitive Sciences*. Nature Macmillan, 2002b.

D. Nikovski. *State-Aggregation Algorithms for Learning Probabilistic Models for Robot Control*. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh,PA, February 2002.

R. Parr. Flexible decomposition algorithms for weakly coupled markov decision problem. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI-98),*, Madison WI, USA, 1998.

S. Parsons and A. Hunter. A review of uncertainty handling formalisms. *Lecture Notes in Computer Science*, 1455:8–37, 1998.

A. Pereira. Programmation bayiésienne de comportements animaux sur un robot mobile. Technical report, Conservatoire Nacional d'Arts et Mï¿½iers (CNAM) Centre d'enseignement de Grenoble, 2005.

P. Pirjanian. Behavior coordination mechanisms - state-of-the-art. Technical Report Tech-report IRIS-99-375, Institute for Robotics and Intelligent Systems, School of Engineering, University of Southern California, Robotics Research Laboratory, University of Southern California, 941 West 37th Place, SAL 203, Los Angeles, CA 90089-0781, USA, 1999.

P. Poupart. *Exploiting Structure to Efficiently Solve Large Scale Partially Observable Markov Decision Processes*. Ph.d. thesis, Department of Computer Science, University of Toronto, Toronto, Canada, 2005.

C. Pradalier. *Navigation intentionnelle d'un robot mobile*. Thèse de doctorat, Inst. Nat. Polytechnique de Grenoble, Grenoble (FR), Sept. 2004.

C. Pradalier, F. Colas, and P. Bessière. Expressing bayesian fusion as a product of distributions: Applications in robotics. In *Proc. of the IEEE-RSJ Int. Conf. on Intelligent Robots and Systems*, Las vegas, NV (US), Oct. 2003a.

C. Pradalier, F. Colas, and P. Bessière. Expressing bayesian fusion as a product of distributions: Application to randomized hough transform. In *Proc. of the Conf. on Bayesian Methods and Maximum Entropy in Science and Engineering*, Jackson Hole, WY (US), Aug. 2003b.

C. Pradalier, J. Hermosillo, C. Koike, C. Braillon, P. Bessière, and C. Laugier. The cycab: a car-like robot navigating autonomously and safely among pedestrians. *Robotics and Autonomous Systems*, 50(1):51–68, 2005.

C. Pradalier, J. Hermosillo, C. Koike, C. Braillon, P. P.Bessière, and C. Laugier. Safe and autonomous navigation for a car-like robot among pedestrian. In *IARP Int. Workshop on Service, Assistive and Personal Robots*, Madrid (ES), Oct. 2003c.

C. Pradalier, J. Hermosillo, K. Koike, C. Braillon, P. Bessière, and C. Laugier. An autonomous car-like robot navigating safely among pedestrians. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, New Orleans, LA (US), Apr. 2004.

T. Prescott, P. Redgrave, and K. Gurney. Layered control architectures in robots and vertebrates. *Adaptive Behavior*, 7:99–127, 1999.

L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–262, 1989.

R. G. Ramirez. *Programmation bayésienne des bras manipulateurs*. Thèse de doctorat, Inst. Nat. Polytechnique de Grenoble, Grenoble (FR), May 2003.

F. Raspail. Apprentissage bayésien par imitation. Mémoire de Diplôme d'Etudes Approfondies, Univ. Joseph Fourier, Grenoble (FR), June 2002.

R. D. Rimey. *Control of Selective Perception using Bayes Nets and Decision Theory*. PhD thesis, Computer Science Dept., U. Rochester, December 1993.

H. Shatkay and L. P. Kaelbling. Learning topological maps with weak local odometric information. In *IJCAI (2)*, pages 920–929, 1997.

R. Simmons and S. Koenig. Probabilistic robot navigation in partially observable environments. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1080–1087, 1995.

E. Simonin. Carte bayésienne et apprentissage. Mémoire de Diplôme d'Etudes Approfondies, Inst. Nat. Polytechnique de Grenoble, June 2004.

c. Soyer, H. Bozma, and Y. İstefanopulos. Attentional sequence-based recognition: Markovian and evidential reasoning. *IEEE Transaction on Systems,Man and Cybernetics - Part B: Cybernetics*, 33(6):937–950, December 2003.

S.Vijayakumar, J.Conradt, T.Shibata, and S.Schaal. Overt visual attention for a humanoid robot. In *Proc. International Conference on Intelligence in Robotics and Autonomous Systems*, pages 2332–2337, Hawaii, 2001.

G. Theocharous and L. P. Kaelbling. Approximate planning in pomdps with macro-actions. Vancouver, december 2003.

S. Thrun. Probabilistic algorithms in robotics. *Artificial inteligence Magazine*, 21(4):93–110, 2000.

S. Thrun. Particle filters in robotics. In *Proceedings of the 17th Annual Conference on Uncertainty in AI (UAI)*, 2002a.

S. Thrun. Robotic mapping: A survey. In G. Lakemeyer and B. Nebel, editors, *Exploring Artificial Intelligence in the New Millenium*. Morgan Kaufmann, 2002b.

S. Thrun, D. Fox, and W. Burgard. Probabilistic mapping of an environment by a mobile robot. In *Proceedings of the IEEE International Conference on Robotics and Automation*, Leuven, Belgium, 1998.

W. Timberlake. Motivational modes in behavior systems. In R. Mowrer and S. Klein, editors, *Handbook of Contemporary Learning Theories*, chapter 5, pages 155–209. Erlbau Associates, Hillsdale, NJ, 2001.

W. Timberlake and G. Lucas. Behavior systems and learning: From misbehavior to general principles. In K. S. and R. Mowrer, editors, *Contemporary Learning Theories: Instrumental Conditioning Theory and the Impact of Biological Constraints*, chapter 9. Lawrence Erlbaum Associates, Inc., Hillsdale NJ, 1989.

W. Timberlake and K. Silva. Appetitive behavior in ethology, psychology, and behavior systems. In N. Thompson, editor, *Perspectives in Ethology, volume 11: Behavioral Design*, chapter 7. Plenum Press, New York, 1995.

X. Tu. *Artificial Animals for Computer Animation: biomechanics, locomotion, perception and behavior*. PhD thesis, University of Toronto, 1996.

T. Tyrrell. *Computational Mechanisms for Action Selection*. PhD thesis, University of Edinburgh, 1993.

C.-J. Westelius. *Focus of Attention and Gaze Control for Robot Vision*. PhD thesis, Linköping University, Sweden, SE-581 83 Linköping, Sweden, 1995. Dissertation No 379, ISBN 91-7871-530-X.

J. D. Williams, P. Poupart, and S. Young. Factored partially observable markov decision processes for dialogue management. In *4th Workshop on Knowledge and Reasoning in Practical Dialog Systems, International Joint Conference on Artificial Intelligence (IJCAI)*, Edinburgh, August 2005.

Part I

Mathematical Details on Prediction-Estimation Loop

Appendix A

Introduction

Appendixes B to E present, in details, mathematical demonstrations for filters presented in section 5 and in chapters 6,7 and 9, respectively.

Even if each chapter presents a different filter, the structure of the mathematical demonstrations is the same, and the same basic rules are employed. It would be unnecessary long to expose the repetitive reasoning and assumptions employed in each appendix, and this appendix groups together common aspects used in the following proofs.

First, the outline used in all demonstrations is presented and discussed in section A.1. Then, some basic mathematical tools employed in these proofs are shown in section A.2.

A.1 Proof Outline

Most important feature of Bayes filter is the recursive calculation loop that allows reutilisation of previous time step calculated expression. In this work, each chapter presented a filter with new features, terms and questions, but it is essential to be sure that recursive calculation is always operative.

Furthermore, in order to reduce necessary computations, elementary filters definition is also important. Utilisation of independent elementary filters can considerably reduce number of calculations, but only if it can be shown that they are equivalent to a global filter including all variables.

In appendixes B to E, two main proofs are shown:

- First, that recursive calculation holds, both for any elementary filter and for a global filter;
- Second, the equivalence between using elementary filters and a global filter is demonstrated.

Filters presented in each chapter have the same structure: joint distributions (global and elementary) and questions (prediction, estimation and motor control questions for

all filters). Some of them include specific questions, as behaviour selection or attention questions. Proofs are then different according to the filter characteristics.

Even if different, recursive calculation and equivalence of utilisation proofs in each appendix share common features and reasoning, presented in the following sections.

A.1.1 Recursive Calculation

At each time step, states prediction and estimation depend on all past time states and this implies that number of calculations would increase with time. However, in a Bayes filter, it is possible to reuse last time step calculations, as demonstrated in chapter 4: states prediction calculation at a time step t uses states estimation at time $t-1$.

Recursive calculation is verified both for elementary filter and for global filter and it consists in the following steps:

- Initially finding the expression for prediction question;
- Then, for all other questions asked to the filter, find the expression in relation to the prediction question;
- Last step consists in rewriting prediction expression in a way that the estimation expression of previous time step can be used, which proves that recursive calculation is possible.

Two important and related aspects are considered: the sequence in which questions are asked and the variables with known values used in right part of each question. Depending on assumptions made relative to these aspects, recursive calculation is not true. Following, the description of the general assumptions employed in this work for all filters.

Questions Sequence

In relation to questions sequence, it is assumed that prediction is always the first question at each time step, and estimation followed by motor question are always the last ones. This assumption is essential to assure that recursive calculation is possible and that motor commands question is based on the most complete states estimation.

All questions eventually asked between prediction and estimation (also called intermediary questions) depend on the filter specific features, and they have to be expressed as function of prediction, assuring reutilisation of calculations already done.

Question sequence is the same in elementary and in global filters: this assumption is necessary for showing the equivalence between using elementary filters and one global filter.

It is assumed that prediction and estimation questions form independent prediction-estimation loops in each elementary filter, because state and observation variables subsets are independent in relation to dynamic and observation model. As a consequence, there is no necessary order of elementary filters when asking these questions.

However, for intermediary questions where coherence based fusion is employed, all filters must have answered the question before fusion can take place. After fusion, a decision is taken regarding values of some internal variables, and these values are necessary to ask next question. Recursive calculation can be affected by intermediary questions sequence, in elementary filters as well as in global filter, but no particular order between elementary filter is considered relevant.

Variables with known values

Sequence of questions is very important regarding knowledge used in the right part of each question (known variables).

For prediction, all variables with known values from previous time steps are present in the right part of question. This means all right part applied for calculating previous time step estimation question, plus motor commands values.

Intermediary questions are asked in the sequence defined by each proposed filter. Right part of each intermediary question consists of:

- All known variables in prediction question;
- Observation variables values, if already available;
- Variables whose values were previously decided.

Whenever coherence-based fusion is applied, the coherence variables are considered to be known equal to true, when the corresponding fusion variable appears at left or at right side of a question.

Estimation is calculated after all intermediary questions are answered and all necessary values are drawn. Observation variables values are always taken into account at estimation question, as well as other variables whose values are decided or known.

Right part of motor question is similar to right part of estimation question, with addition of coherence variables related to motor command variables.

A.1.2 Equivalence between Elementary and Global Filters

When the necessary constraints are fulfilled, the utilisation of elementary filters can be an important asset for reducing complexity and calculation time. However, it is important to verify that the set of elementary filters are equivalent to one global filter. This means showing that using the answers provided by each elementary filter for each asked question is equivalent to using the answers to the same questions asked to a global filter whose joint distribution is defined over all state and observation variables.

We assume for all filters that this equivalence is denoted by a proportionality in the product of elementary answers at a given time step in relation to the global answer. This is demonstrated using mathematical induction.

Each equivalence proof begins by showing that if this proportionality holds for prediction at time step $t-1$, it also holds for estimation at time step $t-1$ and consequently

for prediction at time step t. Then, it is enough to demonstrate that the proportionality holds for prediction at time step $t = 1$, that it is assured that it will be true for prediction and estimation for all following time steps.

Regarding other questions, as motor control and intermediary questions specific for each filter, it is assumed that coherence-based fusion is employed in both elementary filter and global filter joint distributions for defining corresponding variables dependence. From this assumption, it is shown that proportionality also applies for these questions.

A.2 Tools and Rules

When developing the expressions answering the questions, some rules can be employed to simplify the expressions. These rules are described in this section and referenced later in each demonstration.

A.2.1 General Inference Rule

As shown in section 2.1.1, in order to answer a question, the following inference rule is applied:

$$\begin{aligned} &P(Searched|Known\ \pi\ \delta) = \\ &\quad = \sum_{Unknown} \frac{P(SearchedKnownUnknown|\ \pi\ \delta)}{\sum_{Searched,Unknown} P(SearchedKnownUnknown|\ \pi\ \delta))} \end{aligned} \tag{A.1}$$

Actually, expression $\sum_{Searched,Free} P(SearchedKnownFree|\ \pi)$ is considered a normalisation constant in most answer expressions in the form $P(Known|\ \pi)$.

A.2.2 Normalisation

Normalisation rule, as discussed in section 2.1.1, states that when a and b are known values of variables A and B:

$$\sum_X P(X|\ a\ b\ \pi) = 1. \tag{A.2}$$

Normally, more complicated relations happens and we have, for example:

$$\sum_{X,Y} [P(X|\ a\ b\ \pi)P(Y|X\ b\ \pi)] = \sum_X \left[P(X|\ a\ b\ \pi) \sum_Y P(Y|X\ b\ \pi) \right]. \tag{A.3}$$

When summing over Y, values of X are known, due to the outermost sum over X, and then we have:

$$\sum_X \left[P(X|\ a\ b\ \pi) \sum_Y P(Y|X\ b\ \pi) \right] = \sum_X P(X|\ a\ b\ \pi) = 1. \tag{A.4}$$

Another particular situation where normalisation is applied consists in:

$$P(X|\, a\, b\, \pi) \propto P(X|\, a\, \pi) \sum_{Y} P(Y|X\, b\, \pi) = P(X|\, a\, \pi). \tag{A.5}$$

As variable X is located at left in the question, the sum over Y is calculated for each possible value of X: this implies that value of X is known every time the sum is calculated, and the result is one.

Appendix B

Elementary Filters

This appendix deals with Bayes filter introduced in chapter 5. In this section, the idea presented is that elementary filters can be used instead of a global one, and that this utilization actually leads to reduction of computations efforts.

Here, the mathematical details are shown for proving that:

- Elementary filters present recursive calculation loop, in section B.1;
- Global filter defined over all variables also present recursive calculation loop, in section B.2; and
- Utilization of elementary filters is similar utilisation of a global in section B.3.

Reasoning employed in each proof can be found in appendix A, where the general information about the mathematical appendix is presented.

B.1 Elementary Filter

This section aims to prove that in an elementary filter whose joint distribution is defined in section B.1.1, it is possible to take advantage of recursive calculation loop. This proof is organized in four steps: first, expression for prediction is derived. Then, the estimation expression as a function of prediction is found. Following, motor question expression shows that it can be calculated from estimation. Finally, the prediction question is shown to be function of estimation question of previous time step.

B.1.1 Joint Distribution

$$
\begin{aligned}
&P(M^{0:t}S_i^{0:t}Z_i^{0:t}\lambda_i^{0:t}|\pi_i) \\
&= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\ \times P(Z_i^j|S_i^j\pi_i) \\ \times P(M^j|\pi_i) \times P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i) \end{array} \right] \\ \times P(M^0S_i^0Z_i^0\lambda_i^0|\pi_i) \end{array} \right.
\end{aligned}
\tag{B.1}
$$

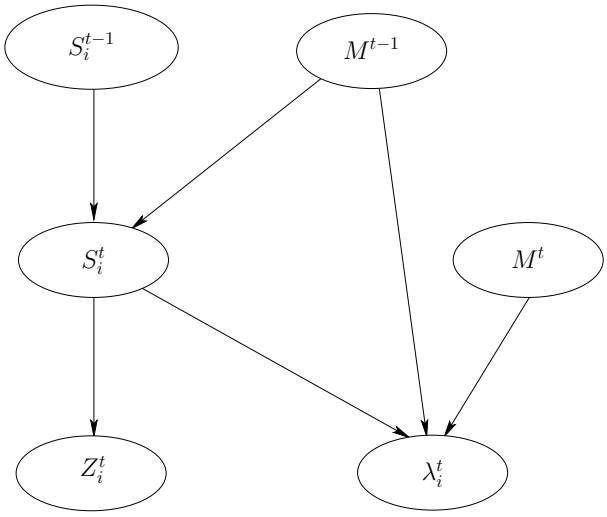


Figure B.1: Bayes Network for Elementary Filter defined in joint distribution equation B.1.

B.1.2 Prediction

States prediction at time instant t is a question asked to the joint distribution, where state variables subset S_i^t is in the left part of the question and all variables with known values at time instant $t-1$ are in the right part: $z_i^{0:t-1}$, $m^{0:t-1}$ and $\lambda_i^{0:t-1}$.

In order to find the expression for prediction question, we employ the rule in equation A.1 and, arranging terms, apply normalization as in equation A.2 in order to simplify the expression.

$$
\begin{aligned}
&P(S_i^t | m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} \pi_i) \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} | \pi_i)} \\ \times \sum_{\left[\begin{array}{l} S_i^{0:t-1} \\ M^t \\ Z_i^t \\ \lambda_i^t \end{array}\right]} \left[\begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{array} \right] \end{array} \right.
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} | \pi_i)} \\ \times \sum \begin{bmatrix} S_i^{0:t-1} \\ M^t \\ Z_i^t \\ \lambda_i^t \end{bmatrix} \begin{bmatrix} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]} Z_i^t | S_i^t \pi_i) \right] \\ \times P(M^t | \pi_i) \times P(\lambda_i^t | M^t S_i^t m^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{bmatrix} \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{bmatrix} \end{array} \right| \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} | \pi_i)} \\ \times \sum_{S_i^{0:t-1}} \begin{bmatrix} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times \sum_{Z_i^t} \left[\prod_{k=1}^{N_{z_i}} \left[P({}_{[k]} Z_i^t | S_i^t \pi_i) \right] \right] \\ \times \sum_{M^t} \left[P(M^t | \pi_i) \times \sum_{\lambda_i^t} P(\lambda_i^t | M^t S_i^t m^{t-1} \pi_i) \right] \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{bmatrix} \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{bmatrix} \end{array} \right| \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} | \pi_i)} \\ \times \sum_{S_i^{0:t-1}} \begin{bmatrix} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{bmatrix} \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{bmatrix} \end{array} \right|
\end{aligned}
\tag{B.2}
$$

B.1.3 Estimation

Estimation question asked to the joint distribution has states at time instant t at left side of question. In the right part, all variables employed in prediction question plus observations at time step t, z_i^t.

The expression for estimation question is found applying the rule in equation A.1, and simplifying with the normalization rule (equation A.2). Expression found is function of prediction question, as desired.

$$
\begin{aligned}
&P(S_i^t | m^{0:t-1} z_i^{0:t} \lambda_i^{0:t-1} \pi_i) \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t} \lambda_i^{0:t-1} | \pi_i)} \\ \times \sum \begin{bmatrix} S_i^{0:t-1} \\ M^t \\ \lambda_i^t \end{bmatrix} \begin{bmatrix} \prod_{j=1}^{t} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{bmatrix} \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{bmatrix} \end{array} \right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t} \lambda_i^{0:t-1} | \pi_i)} \\ \times \sum \begin{bmatrix} S_i^{0:t-1} \\ M^t \\ \lambda_i^t \end{bmatrix} \begin{bmatrix} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]} z_i^t | S_i^t \pi_i) \right] \\ \times P(M^t | \pi_i) \times P(\lambda_i^t | M^t S_i^t m^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{bmatrix} \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{bmatrix} \end{array} \right| \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t} \lambda_i^{0:t-1} | \pi_i)} \\ \times \sum_{S_i^{0:t-1}} \begin{bmatrix} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]} z_i^t | S_i^t \pi_i) \right] \\ \times \sum_{M^t} \left[P(M^t | \pi_i) \times \sum_{\lambda_i^t} P(\lambda_i^t | M^t S_i^t m^{t-1} \pi_i) \right] \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{bmatrix} \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{bmatrix} \end{array} \right| \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t} \lambda_i^{0:t-1} | \pi_i)} \\ \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]} z_i^t | S_i^t \pi_i) \right] \\ \times \sum_{S_i^{0:t-1}} \begin{bmatrix} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{bmatrix} \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{bmatrix} \end{array} \right| \\
&= \left| \begin{array}{l} \frac{P(m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} | \pi_i)}{P(m^{0:t-1} z_i^{0:t} \lambda_i^{0:t-1} | \pi_i)} \\ \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]} z_i^t | S_i^t \pi_i) \right] \\ \times P(S_i^t | m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} \pi_i) \end{array} \right|
\end{aligned} \tag{B.3}
$$

B.1.4 Motor

Left side of motor question contains motor variables at time instant t, and at right side there are all variables at right side of estimation question, plus coherence variables at time instant t, λ_i^t

Motor question is found applying sum rule, in equation A.1, and terms are re-arranged in order to find this expression as a function of estimation question expression.

$$P(M^t | m^{0:t-1} z_i^{0:t} \lambda_i^{0:t} \pi_i)$$

$$
\begin{aligned}
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t} \lambda_i^{0:t} | \pi_i)} \\ \times \sum_{S_i^{0:t}} \left[\begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t} \lambda_i^{0:t} | \pi_i)} \\ \times \sum_{S_i^{0:t}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \prod_{k=1}^{N_{z_i}} \left[P(_{[k]} z_i^t | S_i^t \pi_i) \right] \\ \times P(M^t | \pi_i) \times P(\lambda_i^t | M^t S_i^t m^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t} \lambda_i^{0:t} | \pi_i)} \\ \times \sum_{S_i^t} \left[\begin{array}{l} P(M^t | \pi_i) \times P(\lambda_i^t | M^t S_i^t m^{t-1} \pi_i) \\ \prod_{k=1}^{N_{z_i}} \left[P(_{[k]} z_i^t | S_i^t \pi_i) \right] \\ \times \sum_{S_i^{0:t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(m^{0:t-1} z_i^{0:t} \lambda_i^{0:t-1} | \pi_i)}{P(m^{0:t-1} z_i^{0:t} \lambda_i^{0:t} | \pi_i)} \\ \times \sum_{S_i^t} \left[\begin{array}{l} P(M^t | \pi_i) \times P(\lambda_i^t | M^t S_i^t m^{t-1} \pi_i) \\ \times P(S_i^t | m^{0:t-1} z_i^{0:t} \lambda_i^{0:t-1} \pi_i) \end{array} \right] \end{array} \right.
\end{aligned}
\tag{B.4}
$$

B.1.5 Prediction using estimation

In order to close calculation loop, it is still necessary to show that from estimation at previous time step, it is possible to calculate prediction at present time step. The sum rule, in equation A.1, is initially applied and then terms in equation are rearranged.

$$
\begin{aligned}
&P(S_i^t | m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} \pi_i) \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} | \pi_i)} \\ \times \sum_{S_i^{0:t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{array} \right] \end{array} \right.
\end{aligned}
$$

$$
= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} | \pi_i)} \\ \times \sum_{S_i^{0:t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(S_i^{t-1} | S_i^{t-2} m^{t-2} \pi_i) \\ \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]} z_i^{t-1} | S_i^{t-1} \pi_i) \right] \\ \times P(m^{t-1} | \pi_i) \times P(\lambda_i^{t-1} | m^{t-1} S_i^{t-1} m^{t-2} \pi_i) \\ \times \prod_{j=1}^{t-2} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{array} \right] \end{array} \right|
$$

$$
= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} | \pi_i)} \\ \times \sum_{S_i^{t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(m^{t-1} | \pi_i) \times P(\lambda_i^{t-1} | m^{t-1} S_i^{t-1} m^{t-2} \pi_i) \\ \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]} z_i^{t-1} | S_i^{t-1} \pi_i) \right] \\ \times \sum_{S_i^{0:t-2}} \left[\begin{array}{l} P(S_i^{t-1} | S_i^{t-2} m^{t-2} \pi_i) \\ \times \prod_{j=1}^{t-2} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 \lambda_i^0 | \pi_i) \end{array} \right] \end{array} \right] \end{array} \right|
$$

$$
= \left| \begin{array}{l} \frac{P(m^{0:t-2} z_i^{0:t-1} \lambda_i^{0:t-2} | \pi_i)}{P(m^{0:t-2} z_i^{0:t-2} \lambda_i^{0:t-2} | \pi_i) \times P(m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} | \pi_i)} \\ \times \sum_{S_i^{t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(m^{t-1} | \pi_i) \times P(\lambda_i^{t-1} | m^{t-1} S_i^{t-1} m^{t-2} \pi_i) \\ \times P(S_i^{t-1} | m^{0:t-2} z_i^{0:t-1} \lambda_i^{0:t-2} \pi_i) \end{array} \right] \end{array} \right| \tag{B.5}
$$

B.2 Global Filter

This section intends to proof that a global filter composed of elementary filters (defined as shown in section B.2.1) also presents a calculation loop. It is organized as the previous section, and follows the same sequence.

B.2.1 Joint Distribution

$$
P(M^{0:t} S^{0:t} Z^{0:t} \lambda^{0:t} | \pi_{lf}) \tag{B.6}
$$

$$
= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \right] \\ \prod_{i=1}^{N_i} \left[P(Z_i^j | S_i^j \pi_i) \right] \\ \times P(M^j | \pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \right] \end{array} \right] \\ \times P(M^0 S^0 Z^0 \lambda^0 | \pi_{lf}) \end{array} \right| \tag{B.7}
$$

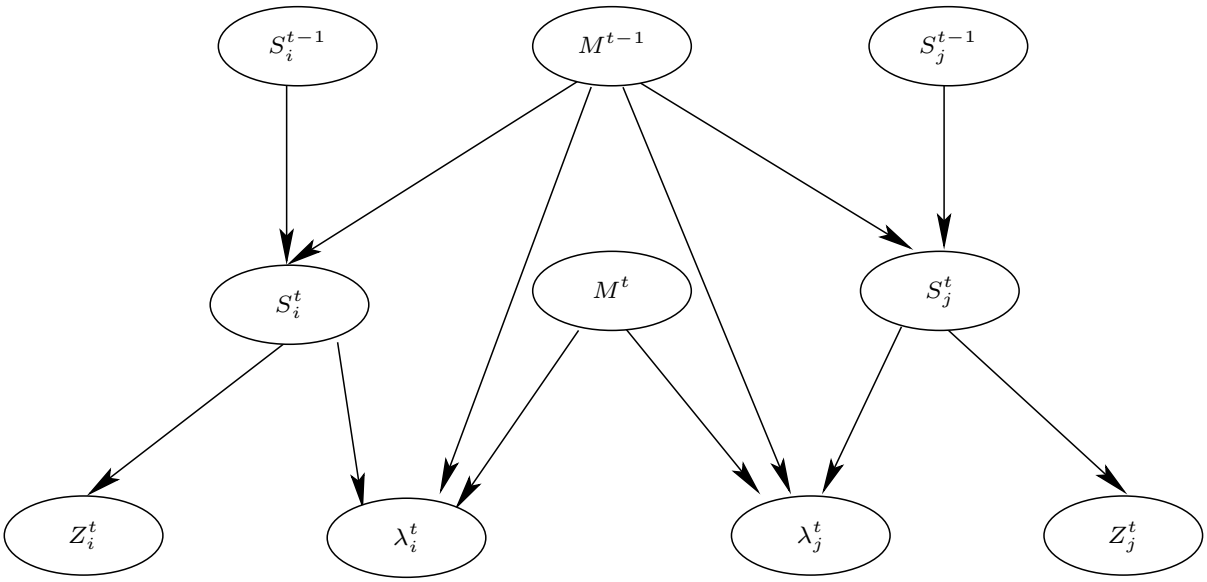


Figure B.2: Bayes Network for Global Filter in joint distribution in equation B.7.

B.2.2 Prediction

$$
\begin{aligned}
&P(S^t|m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}\pi_{lf}) \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})} \\ \times \sum \begin{bmatrix} S^{0:t-1} \\ M^t \\ Z^t \\ \lambda^t \end{bmatrix} \begin{bmatrix} \prod_{j=1}^{t} \begin{bmatrix} \prod_{i=1}^{N_i} \left[P(S_i^j|S_i^{j-1}M^{j-1}\pi_i)\right] \\ \prod_{i=1}^{N_i} \left[P(Z_i^j|S_i^j\pi_i)\right] \\ \times P(M^j|\pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i)\right] \end{bmatrix} \\ \times P(M^0S^0Z^0\lambda^0|\pi_{lf}) \end{bmatrix} \end{array} \right. \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})} \\ \times \sum \begin{bmatrix} S^{0:t-1} \\ M^t \\ Z^t \\ \lambda^t \end{bmatrix} \begin{bmatrix} \prod_{i=1}^{N_i} \left[P(S_i^t|S_i^{t-1}M^{t-1}\pi_i)\right] \\ \prod_{i=1}^{N_i} \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]}Z_i^t|S_i^t\pi_i)\right] \\ \times P(M^t|\pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^t|M^tS_i^tm^{t-1}\pi_i)\right] \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} \prod_{i=1}^{N_i} \left[P(S_i^j|S_i^{j-1}M^{j-1}\pi_i)\right] \\ \prod_{i=1}^{N_i} \left[P(Z_i^j|S_i^j\pi_i)\right] \\ \times P(M^j|\pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i)\right] \end{bmatrix} \\ \times P(M^0S^0Z^0\lambda^0|\pi_{lf}) \end{bmatrix} \end{array} \right.
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})} \\ \times \sum_{S^{0:t-1}} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^t|S_i^{t-1}M^{t-1}\pi_i) \right] \\ \times \sum_{Z^t} \left[\prod_{i=1}^{N_i} \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]}Z_i^t|S_i^t\pi_i) \right] \right] \\ \times \sum_{M^t} \left[P(M^t|\pi_i) \times \sum_{\lambda^t} \prod_{i=1}^{N_i} \left[P(\lambda_i^t|M^tS_i^tm^{t-1}\pi_i) \right] \right] \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \left[P(Z_i^j|S_i^j\pi_i) \right] \\ \times P(M^j|\pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i) \right] \end{array} \right] \\ \times P(M^0S^0Z^0\lambda^0|\pi_{lf}) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})} \\ \times \sum_{S^{0:t-1}} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^t|S_i^{t-1}M^{t-1}\pi_i) \right] \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \left[P(Z_i^j|S_i^j\pi_i) \right] \\ \times P(M^j|\pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i) \right] \end{array} \right] \\ \times P(M^0S^0Z^0\lambda^0|\pi_{lf}) \end{array} \right] \end{array} \right.
\end{aligned}
\tag{B.8}
$$

B.2.3 Estimation

$$
\begin{aligned}
&P(S^t|m^{0:t-1}z^{0:t}\lambda^{0:t-1}\pi_{lf}) \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t}\lambda^{0:t-1}|\pi_{lf})} \\ \times \sum \left[\begin{array}{l} S^{0:t-1} \\ M^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \left[P(Z_i^j|S_i^j\pi_i) \right] \\ \times P(M^j|\pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i) \right] \end{array} \right] \\ \times P(M^0S^0Z^0\lambda^0|\pi_{lf}) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t}\lambda^{0:t-1}|\pi_{lf})} \\ \times \sum \left[\begin{array}{l} S^{0:t-1} \\ M^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^t|S_i^{t-1}M^{t-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]}z_i^t|S_i^t\pi_i) \right] \\ \times P(M^t|\pi_i) \prod_{i=1}^{N_i} \left[P(\lambda_i^t|M^tS_i^tm^{t-1}\pi_i) \right] \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \left[P(Z_i^j|S_i^j\pi_i) \right] \\ \times P(M^j|\pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i) \right] \end{array} \right] \\ \times P(M^0S^0Z^0\lambda^0|\pi_{lf}) \end{array} \right] \end{array} \right.
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t}\lambda^{0:t-1}|\pi_{lf})} \\ \times \sum_{S^{0:t-1}} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^t|S_i^{t-1}M^{t-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]}z_i^t|S_i^t\pi_i) \right] \\ \times \sum_{M^t} \left[P(M^t|\pi_i) \times \sum_{\lambda^t} \prod_{i=1}^{N_i} \left[P(\lambda_i^t|M^tS_i^tm^{t-1}\pi_i) \right] \right] \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \left[P(Z_i^j|S_i^j\pi_i) \right] \\ \times P(M^j|\pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i) \right] \end{array} \right] \\ \times P(M^0S^0Z^0\lambda^0|\pi_{lf}) \end{array} \right] \end{array} \right| \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t}\lambda^{0:t-1}|\pi_{lf})} \\ \prod_{i=1}^{N_i} \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]}z_i^t|S_i^t\pi_i) \right] \\ \times \sum_{S^{0:t-1}} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^t|S_i^{t-1}M^{t-1}\pi_i) \right] \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \left[P(Z_i^j|S_i^j\pi_i) \right] \\ \times P(M^j|\pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i) \right] \end{array} \right] \\ \times P(M^0S^0Z^0\lambda^0|\pi_{lf}) \end{array} \right] \end{array} \right| \\
&= \left| \begin{array}{l} \frac{P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})}{P(m^{0:t-1}z^{0:t}\lambda^{0:t-1}|\pi_{lf})} \\ \prod_{i=1}^{N_i} \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]}z_i^t|S_i^t\pi_i) \right] \times P(S^t|z^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\pi_{lf}) \end{array} \right.
\end{aligned}
\tag{B.9}
$$

B.2.4 Motor Commands

$$
\begin{aligned}
&P(M^t|m^{0:t-1}z^{0:t}\lambda^{0:t}\pi_{lf}) \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t}\lambda^{0:t}|\pi_{lf})} \\ \times \sum_{S^{0:t}} \left[\begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \left[P(Z_i^j|S_i^j\pi_i) \right] \\ \times P(M^j|\pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i) \right] \end{array} \right] \\ \times P(M^0S^0Z^0\lambda^0|\pi_{lf}) \end{array} \right] \end{array} \right| \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t}\lambda^{0:t}|\pi_{lf})} \\ \times \sum_{S^{0:t}} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^t|S_i^{t-1}M^{t-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]}z_i^t|S_i^t\pi_i) \right] \\ \times P(M^t|\pi_i) \prod_{i=1}^{N_i} \left[P(\lambda_i^t|M^tS_i^tm^{t-1}\pi_i) \right] \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \left[P(Z_i^j|S_i^j\pi_i) \right] \\ \times P(M^j|\pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^jS_i^jM^{j-1}\pi_i) \right] \end{array} \right] \\ \times P(M^0S^0Z^0\lambda^0|\pi_{lf}) \end{array} \right] \end{array} \right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(m^{0:t-1}z^{0:t}\lambda^{0:t}|\pi_{lf})} \\
\times \sum_{S^t} \left[\begin{array}{l}
P(M^t|\pi_i) \prod_{i=1}^{N_i} \left[P(\lambda_i^t | M^t S_i^t m^{t-1} \pi_i) \right] \\
\prod_{i=1}^{N_i} \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]}z_i^t | S_i^t \pi_i) \right] \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(S_i^t | S_i^{t-1} M^{t-1} \pi_i) \right] \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \right] \\
\prod_{i=1}^{N_i} \left[P(Z_i^j | S_i^j \pi_i) \right] \\
\times P(M^j | \pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \right]
\end{array} \right] \\
\times P(M^0 S^0 Z^0 \lambda^0 | \pi_{lf})
\end{array} \right]
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(m^{0:t-1}z^{0:t}\lambda^{0:t-1}|\pi_{lf})}{P(m^{0:t-1}z^{0:t}\lambda^{0:t}|\pi_{lf})} \times \sum_{S^t} \left[\begin{array}{l}
P(M^t|\pi_i) \prod_{i=1}^{N_i} \left[P(\lambda_i^t | M^t S_i^t m^{t-1} \pi_i) \right] \\
\times P(S^t | m^{0:t-1} z^{0:t} \lambda^{0:t-1} \pi_{lf})
\end{array} \right]
\end{array} \right|
\end{aligned}
\tag{B.10}
$$

B.2.5 Prediction using Estimation

$$
\begin{aligned}
&P(S^t | m^{0:t-1} z^{0:t-1} \lambda^{0:t-1} \pi_{lf}) \\
&= \left| \begin{array}{l}
\frac{1}{P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})} \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \right] \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \right] \\
\prod_{i=1}^{N_i} \left[P(Z_i^j | S_i^j \pi_i) \right] \\
\times P(M^j | \pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \right]
\end{array} \right] \\
\times P(M^0 S^0 Z^0 \lambda^0 | \pi_{lf})
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})} \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \right] \\
\prod_{i=1}^{N_i} \left[P(S_i^{t-1} | S_i^{t-2} m^{t-2} \pi_i) \right] \\
\prod_{i=1}^{N_i} \prod_{k=1}^{N_{z_i}} \left[P({}_{[k]}z_i^{t-1} | S_i^{t-1} \pi_i) \right] \\
\times P(m^{t-1} | \pi_i) \prod_{i=1}^{N_i} \left[P(\lambda_i^{t-1} | m^{t-1} S_i^{t-1} m^{t-2} \pi_i) \right] \\
\times \prod_{j=1}^{t-2} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \right] \\
\prod_{i=1}^{N_i} \left[P(Z_i^j | S_i^j \pi_i) \right] \\
\times P(M^j | \pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \right]
\end{array} \right] \\
\times P(M^0 S^0 Z^0 \lambda^0 | \pi_{lf})
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})} \\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \right] \\
\times P(m^{t-1}|\pi_i) \prod_{i=1}^{N_i} \left[P(\lambda_i^{t-1} | m^{t-1} S_i^{t-1} m^{t-2} \pi_i) \right] \\
\prod_{i=1}^{N_i} \prod_{k=1}^{N_{z_i}} \left[P(_{[k]} z_i^{t-1} | S_i^{t-1} \pi_i) \right] \\
\times \sum_{S^{0:t-2}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(S_i^{t-1} | S_i^{t-2} m^{t-2} \pi_i) \right] \\
\times \prod_{j=1}^{t-2} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \right] \\
\prod_{i=1}^{N_i} \left[P(Z_i^j | S_i^j \pi_i) \right] \\
\times P(M^j | \pi_{lf}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j | M^j S_i^j M^{j-1} \pi_i) \right]
\end{array} \right] \\
\times P(M^0 S^0 Z^0 \lambda^0 | \pi_{lf})
\end{array} \right]
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(m^{0:t-2}z^{0:t-1}\lambda^{0:t-2}|\pi_{lf})}{P(m^{0:t-2}z^{0:t-2}\lambda^{0:t-2}|\pi_{lf}) \times P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})} \\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
P(m^{t-1}|\pi_{lf}) \prod_{i=1}^{N_i} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(\lambda_i^{t-1} | m^{t-1} S_i^{t-1} m^{t-2} \pi_i)
\end{array} \right] \\
\times P(S^{t-1} | m^{0:t-2} z^{0:t-1} \lambda^{0:t-2} \pi_{lf})
\end{array} \right]
\end{array} \right| \tag{B.11}
$$

B.3 Relations Between Elementary and Global Filter

We would like to demonstrate that global filter prediction at arbitrary time step t is proportional to the product of elementary filter predictions at the same time step, as shown in the following equation:

$$
\begin{aligned}
& P(S^t | m^{0:t-1} z^{0:t-1} \lambda^{0:t-1} \pi_{lf}) \\
& \quad = \kappa_{t-1} \prod_{i=1}^{N_i} \left[P(S_i^t | m^{0:t-1} z_i^{0:t-1} \lambda_i^{0:t-1} \pi_i) \right],
\end{aligned} \tag{B.12}
$$

where κ_{t-1} is a constant which depends on known values of $m^{0:t-1}$, $z^{0:t-1}$ and $\lambda_i^{0:t-1}$. We begin by assuming that this equality holds at time step $t-1$:

$$
\begin{aligned}
& P(S^{t-1} | m^{0:t-2} z^{0:t-2} \lambda^{0:t-2} \pi_{lf}) \\
& \quad = \kappa_{t-2} \prod_{i=1}^{N_i} \left[P(S_i^{t-1} | m^{0:t-2} z_i^{0:t-2} \lambda_i^{0:t-2} \pi_i) \right].
\end{aligned} \tag{B.13}
$$

So, estimation of global filter at times step $t-1$ is found to be:

$$
\begin{aligned}
& P(S^{t-1} | m^{0:t-2} z^{0:t-1} \lambda^{0:t-2} \pi_{lf}) \\
& = \left| \begin{array}{l}
\frac{P(m^{0:t-2}z^{0:t-2}\lambda^{0:t-2}|\pi_{lf})}{P(m^{0:t-2}z^{0:t-1}\lambda^{0:t-2}|\pi_{lf})} \\
\prod_{i=1}^{N_i} \prod_{k=1}^{N_{z_i}} \left[P(_{[k]} z_i^{t-1} | S_i^{t-1} \pi_i) \right] \\
\times P(S^{t-1} | z^{0:t-2} m^{0:t-2} \lambda^{0:t-2} \pi_{lf})
\end{array} \right.
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l} \frac{P(m^{0:t-2}z^{0:t-2}\lambda^{0:t-2}|\pi_{lf})}{P(m^{0:t-2}z^{0:t-1}\lambda^{0:t-2}|\pi_{lf})} \\ \prod_{i=1}^{N_i}\prod_{k=1}^{N_{z_i}}\left[P({}_{[k]}z_i^{t-1}|S_i^{t-1}\pi_i)\right] \\ \times\kappa_{t-2}\prod_{i=1}^{N_i}\left[P(S_i^{t-1}|m^{0:t-2}z_i^{0:t-2}\lambda_i^{0:t-2}\pi_i)\right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(m^{0:t-2}z^{0:t-2}\lambda^{0:t-2}|\pi_{lf})}{P(m^{0:t-2}z^{0:t-1}\lambda^{0:t-2}|\pi_{lf})} \times \kappa_{t-2} \\ \prod_{i=1}^{N_i}\left[\begin{array}{l} \prod_{k=1}^{N_{z_i}}\left[P({}_{[k]}z_i^{t-1}|S_i^{t-1}\pi_i)\right] \\ \times P(S_i^{t-1}|m^{0:t-2}z_i^{0:t-2}\lambda_i^{0:t-2}\pi_i) \end{array}\right] \end{array} \right.
\end{aligned}
\tag{B.14}
$$

From the equation above and equation B.3, we can find:

$$
\begin{aligned}
&P(S^{t-1}|m^{0:t-2}z^{0:t-1}\lambda^{0:t-2}\pi_{lf}) \\
&= \left| \begin{array}{l} \frac{P(m^{0:t-2}z^{0:t-2}\lambda^{0:t-2}|\pi_{lf})}{P(m^{0:t-2}z^{0:t-1}\lambda^{0:t-2}|\pi_{lf})} \times \kappa_{t-2}\prod_{i=1}^{N_i}\left[\frac{P(m^{0:t-2}z_i^{0:t-2}\lambda_i^{0:t-2}|\pi_i)}{P(m^{0:t-2}z_i^{0:t-1}\lambda_i^{0:t-2}|\pi_i)}\right] \\ \prod_{i=1}^{N_i}\left[\ P(S_i^{t-1}|m^{0:t-2}z_i^{0:t-1}\lambda_i^{0:t-2}\pi_i)\ \right] \end{array} \right.
\end{aligned}
\tag{B.15}
$$

Until now, we have seen that assuring prediction in global filter to be proportional to the product of predictions of elementary filters, we can be sure that the same holds for estimation. Now, we will verify that this implies also propotionality for prediction at time t.

From estimation equation above, prediction at global filter in time step t can then be found to be:

$$
\begin{aligned}
&P(S^t|m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}\pi_{lf}) \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})} \\ \times\sum_{S^{t-1}}\left[\begin{array}{l} P(m^{t-1}|\pi_{lf}) \\ \prod_{i=1}^{N_i}\left[\begin{array}{l} P(\lambda_i^{t-1}|m^{t-1}S_i^{t-1}m^{t-2}\pi_i) \\ P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \end{array}\right] \\ \times P(S^{t-1}|m^{0:t-2}z^{0:t-1}\lambda^{0:t-2}\pi_{lf}) \end{array}\right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})} \\ \times\sum_{S^{t-1}}\left[\begin{array}{l} P(m^{t-1}|\pi_{lf}) \\ \prod_{i=1}^{N_i}\left[\begin{array}{l} P(\lambda_i^{t-1}|m^{t-1}S_i^{t-1}m^{t-2}\pi_i) \\ P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \end{array}\right] \\ \times\frac{P(m^{0:t-2}z^{0:t-2}\lambda^{0:t-2}|\pi_{lf})}{P(m^{0:t-2}z^{0:t-1}\lambda^{0:t-2}|\pi_{lf})} \times \kappa_{t-2}\prod_{i=1}^{N_i}\left[\frac{P(m^{0:t-2}z_i^{0:t-2}\lambda_i^{0:t-2}|\pi_i)}{P(m^{0:t-2}z_i^{0:t-1}\lambda_i^{0:t-2}|\pi_i)}\right] \\ \prod_{i=1}^{N_i}\left[\ P(S_i^{t-1}|m^{0:t-2}z_i^{0:t-1}\lambda_i^{0:t-2}\pi_i)\ \right] \end{array}\right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})} \times \frac{P(m^{0:t-2}z^{0:t-2}\lambda^{0:t-2}|\pi_{lf})}{P(m^{0:t-2}z^{0:t-1}\lambda^{0:t-2}|\pi_{lf})} \times \kappa_{t-2}\prod_{i=1}^{N_i}\left[\frac{P(m^{0:t-2}z_i^{0:t-2}\lambda_i^{0:t-2}|\pi_i)}{P(m^{0:t-2}z_i^{0:t-1}\lambda_i^{0:t-2}|\pi_i)}\right] \\ \times\frac{P(m^{t-1}|\pi_{lf})}{\prod_{i=1}^{N_i}P(m^{t-1}|\pi_i)} \\ \times\sum_{S^{t-1}}\left[\begin{array}{l} \prod_{i=1}^{N_i}P(m^{t-1}|\pi_i) \\ \prod_{i=1}^{N_i}\left[\begin{array}{l} P(\lambda_i^{t-1}|m^{t-1}S_i^{t-1}m^{t-2}\pi_i) \\ P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \end{array}\right] \\ \prod_{i=1}^{N_i}\left[\ P(S_i^{t-1}|m^{0:t-2}z_i^{0:t-1}\lambda_i^{0:t-2}\pi_i)\ \right] \end{array}\right] \end{array} \right.
\end{aligned}
\tag{B.16}
$$

Comparing the equation above with equation B.5, we can see that:

$$
\begin{aligned}
&P(S^t|m^{0:t-1}z^{0:t-1}\lambda_i^{0:t-1}\pi_{lf}) \\
&= \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})} \times \frac{P(m^{0:t-2}z^{0:t-2}\lambda^{0:t-2}|\pi_{lf})}{P(m^{0:t-2}z^{0:t-1}\lambda^{0:t-2}|\pi_{lf})} \times \kappa_{t-2} \\ \times \frac{P(m^{t-1}|\pi_{lf})}{\prod_{i=1}^{N_i} P(m^{t-1}|\pi_i)} \\ \prod_{i=1}^{N_i} \left[\frac{P(m^{0:t-2}z_i^{0:t-1}\lambda_i^{0:t-2}|\pi_i)}{P(m^{0:t-2}z_i^{0:t-2}\lambda_i^{0:t-2}|\pi_i) \times P(m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}|\pi_i)} \right] \\ \prod_{i=1}^{N_i} P(S_i^t|m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}\pi_i) \end{array} \right.
\end{aligned}
\tag{B.17}
$$

We can conclude that given proportionality at time $t-1$ as initial assumption in equation B.13, it will hold at time step t, as shown above in equation B.17. Proportionality constant κ can then be calculated as:

$$
\kappa_{t-1} = \left| \begin{array}{l} \frac{1}{P(m^{0:t-1}z^{0:t-1}\lambda^{0:t-1}|\pi_{lf})} \times \frac{P(m^{0:t-2}z^{0:t-2}\lambda^{0:t-2}|\pi_{lf})}{P(m^{0:t-2}z^{0:t-1}\lambda^{0:t-2}|\pi_{lf})} \times \kappa_{t-2} \\ \times \frac{P(m^{t-1}|\pi_{lf})}{\prod_{i=1}^{N_i} P(m^{t-1}|\pi_i)} \\ \prod_{i=1}^{N_i} \left[\frac{P(m^{0:t-2}z_i^{0:t-1}\lambda_i^{0:t-2}|\pi_i)}{P(m^{0:t-2}z_i^{0:t-2}\lambda_i^{0:t-2}|\pi_i) \times P(m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}|\pi_i)} \right] \end{array} \right.
\tag{B.18}
$$

Consequently, if it is possible to demonstrate that this assumption is true for $t = 1$, it will be true for all sucessive time steps.

Lets first write global prediction expression for $t = 1$:

$$
\begin{aligned}
&P(S^1|m^0z^0\lambda^0\pi_{lf}) \\
&= \left| \begin{array}{l} \frac{1}{P(m^0z^0\lambda^0|\pi_{lf})} \\ \times \sum \begin{bmatrix} S^0 \\ Z^1 \\ M^1 \\ \lambda^1 \end{bmatrix} \begin{bmatrix} P(m^0|\pi_{lf}) \times P(S^0|\pi_{lf}) \times P(z^0|S^0\pi_{lf}) \times P(\lambda^0|m^0S^0\pi_{lf}) \\ \prod_{i=1}^{N_i} P(S_i^1|S_i^0m^0\pi_i) \\ \prod_{i=1}^{N_i} P(Z_i^1|S_i^1\pi_i) \\ \times P(M^1|\pi_{lf}) \\ \prod_{i=1}^{N_i} [P(\lambda_i^1|M^1S_i^1m^0\pi_i)] \end{bmatrix} \end{array} \right.
\end{aligned}
\tag{B.19}
$$

$$
= \left| \begin{array}{l} \frac{1}{P(m^0z^0\lambda^0|\pi_{lf})} \\ \times \sum_{Z^1} \prod_{i=1}^{N_i} P(Z_i^1|S_i^1\pi_{lf}) \\ \times \sum_{M^1} P(M^1|\pi_{lf}) \times \sum_{\lambda^1} \prod_{i=1}^{N_i} [P(\lambda_i^1|M^1S_i^1m^0\pi_{lf})] \\ \times \sum_{S^0} \begin{bmatrix} P(m^0|\pi_{lf}) \times P(S^0|\pi_{lf}) \times P(z^0|S^0\pi_{lf}) \times P(\lambda^0|m^0S^0\pi_{lf}) \\ \prod_{i=1}^{N_i} P(S_i^1|S_i^0m^0\pi_i) \end{bmatrix} \end{array} \right.
\tag{B.20}
$$

$$
= \left| \begin{array}{l} \frac{P(m^0|\pi_{lf})}{P(m^0z^0\lambda^0|\pi_{lf})} \\ \times \sum_{S^0} \left[\prod_{i=1}^{N_i} \begin{bmatrix} P(S_i^0|\pi_i) \times P(z_i^0|S_i^0\pi_i) \times P(\lambda_i^0|m^0S_i^0\pi_i) \\ \times P(S_i^1|S_i^0m^0\pi_i) \end{bmatrix} \right] \end{array} \right.
\tag{B.21}
$$

$$= \left| \begin{array}{l} \frac{P(m^0|\pi_{lf})}{P(m^0 z^0 \lambda^0|\pi_{lf})} \\ \prod_{i=1}^{N_i} \left[\sum_{S_i^0} \left[\begin{array}{l} P(S_i^0|\pi_{lf}) \times P(z_i^0|S_i^0\pi_i) \times P(\lambda_i^0|m^0 S_i^0 \pi_i) \\ \times P(S_i^1|S_i^0 m^0 \pi_i) \end{array} \right] \right] \end{array} \right. \tag{B.22}$$

As elementary filter prediction expression at time step $t = 1$ is:

$$P(S_i^1|m^0 z_i^0 \lambda_i^0 \pi_i)$$

$$= \left| \begin{array}{l} \frac{P(m^0|\pi_i)}{P(m^0 z_i^0 \lambda_i^0|\pi_i)} \\ \times \sum \left[\begin{array}{l} S_i^0 \\ Z_i^1 \\ M^1 \\ \lambda_i^1 \end{array} \right] \left[\begin{array}{l} P(S_i^0|\pi_i) \times P(z_i^0|S_i^0\pi_i) \\ \times P(\lambda_i^0|m^0 S_i^0 \pi_i) \times P(S_i^1|S_i^0 m^0 \pi_i) \\ \times P(Z_i^1|S_i^1 \pi_i) \\ \times P(M^1|\pi_i) \times P(\lambda_i^1|M^1 S_i^1 m^0 \pi_i) \end{array} \right] \end{array} \right. \tag{B.23}$$

$$= \left| \begin{array}{l} \frac{P(m^0|\pi_i)}{P(m^0 z_i^0 \lambda_i^0|\pi_i)} \\ \times \sum_{Z_i^1} P(Z_i^1|S_i^1\pi_i) \times \sum_{M^1} \left[P(M^1|\pi_i) \times \sum_{\lambda_i^1} P(\lambda_i^1|M^1 S_i^1 m^0 \pi_i) \right] \\ \times \sum_{S_i^0} \left[\begin{array}{l} P(S_i^0|\pi_i) \times P(z_i^0|S_i^0\pi_i) \\ \times P(\lambda_i^0|m^0 S_i^0 \pi_i) \times P(S_i^1|S_i^0 m^0 \pi_i) \end{array} \right] \end{array} \right. \tag{B.24}$$

$$= \left| \frac{P(m^0|\pi_i)}{P(m^0 z_i^0 \lambda_i^0|\pi_i)} \times \sum_{S_i^0} \left[\begin{array}{l} P(S_i^0|\pi_i) \times P(z_i^0|S_i^0\pi_i) \\ \times P(\lambda_i^0|m^0 S_i^0 \pi_i) \times P(S_i^1|S_i^0 m^0 \pi_i) \end{array} \right] \right. \tag{B.25}$$

We come to the following relation between prediction in global filter and prediction in elementary filters:

$$P(S^1|m^0 z^0 \lambda^0 \pi_{lf})$$

$$= \left| \begin{array}{l} \frac{P(m^0|\pi_{lf})}{P(m^0 z^0 \lambda^0|\pi_{lf})} \\ \prod_{i=1}^{N_i} \left[\frac{P(m^0 z_i^0 \lambda_i^0|\pi_i)}{P(m^0|\pi_i)} \times P(S_i^1|m^0 z_i^0 \lambda_i^0 \pi_i) \right] \end{array} \right. \tag{B.26}$$

$$= \left| \begin{array}{l} \frac{P(m^0|\pi_{lf})}{P(m^0 z^0 \lambda^0|\pi_{lf})} \prod_{i=1}^{N_i} \left[\frac{P(m^0 z_i^0 \lambda_i^0|\pi_i)}{P(m^0|\pi_i)} \right] \\ \prod_{i=1}^{N_i} \left[P(S_i^1|m^0 z_i^0 \lambda_i^0 \pi_i) \right] \end{array} \right. \tag{B.27}$$

This relation shows that proportionality exists at $t = 1$, and as shown in equation B.17, at $t = 2, 3, \ldots$. In addition, the equation above establishes how to calculate the constant κ_0:

$$\kappa_0 = \frac{P(m^0|\pi_{lf})}{P(m^0 z^0 \lambda^0|\pi_{lf})} \prod_{i=1}^{N_i} \left[\frac{P(m^0 z_i^0 \lambda_i^0|\pi_i)}{P(m^0|\pi_i)} \right] \tag{B.28}$$

Values of κ for all successive times steps can then be calculated as given by equation B.18.

As for motor commands, we can see that:

$$
\begin{aligned}
& P(M^t|m^{0:t-1}z^{0:t}\lambda^{0:t}\pi_{lf}) \\
&= \sum_{S^t} \left[\begin{array}{l} P(M^t|\pi_{lf}) \prod_{i=1}^{N_i} [P(\lambda_i^t|M^t S_i^t m^{t-1}\pi_i)] \\ \times P(S^t|m^{0:t-1}z^{0:t}\lambda^{0:t-1}\pi_{lf}) \end{array} \right] \\
&= \sum_{S^t} \left[\begin{array}{l} P(M^t|\pi_{lf}) \prod_{i=1}^{N_i} [P(\lambda_i^t|M^t S_i^t m^{t-1}\pi_i)] \\ \times \frac{P(m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}|\pi_i)}{P(m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}|\pi_i)} \times \kappa_{t-1} \prod_{i=1}^{N_i} \left[\frac{P(m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}|\pi_i)}{P(m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}|\pi_i)} \right] \\ \prod_{i=1}^{N_i} \left[\; P(S_i^t|m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}\pi_i) \; \right] \end{array} \right]
\end{aligned}
\tag{B.29}
$$

Considering that all constants can be taken out of the state sum, we have:

$$
\begin{aligned}
& P(M^t|m^{0:t-1}z^{0:t}\lambda^{0:t}\pi_{lf}) \\
&= \left| \begin{array}{l} \frac{P(m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}|\pi_i)}{P(m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}|\pi_i)} \times \kappa_{t-1} \prod_{i=1}^{N_i} \left[\frac{P(m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}|\pi_i)}{P(m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}|\pi_i)} \right] \\ \times \sum_{S^t} \left[\begin{array}{l} P(M^t|\pi_{lf}) \prod_{i=1}^{N_i} [P(\lambda_i^t|M^t S_i^t m^{t-1}\pi_i)] \\ \prod_{i=1}^{N_i} \left[\; P(S_i^t|m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}\pi_i) \; \right] \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}|\pi_i)}{P(m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}|\pi_i)} \times \kappa_{t-1} \prod_{i=1}^{N_i} \left[\frac{P(m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}|\pi_i)}{P(m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}|\pi_i)} \right] \\ \times \frac{P(M^t|\pi_{lf})}{\prod_{i=1}^{N_i} P(M^t|\pi_i)} \\ \prod_{i=1}^{N_i} \sum_{S_i^t} \left[\begin{array}{l} P(M^t|\pi_i) \times P(\lambda_i^t|M^t S_i^t m^{t-1}\pi_i) \\ \times P(S_i^{t-1}|m^{0:t-2}z_i^{0:t-1}\lambda_i^{0:t-2}\pi_i) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}|\pi_i)}{P(m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}|\pi_i)} \times \kappa_{t-1} \prod_{i=1}^{N_i} \left[\frac{P(m^{0:t-1}z_i^{0:t-1}\lambda_i^{0:t-1}|\pi_i)}{P(m^{0:t-1}z_i^{0:t}\lambda_i^{0:t-1}|\pi_i)} \right] \\ \times \frac{P(M^t|\pi_{lf})}{\prod_{i=1}^{N_i} P(M^t|\pi_i)} \prod_{i=1}^{N_i} [P(m^{0:t-1}z_i^{0:t}\lambda_i^{0:t}|\pi_i)] \\ \prod_{i=1}^{N_i} P(M^t|m^{0:t-1}z_i^{0:t}\lambda_i^{0:t}\pi_i) \end{array} \right.
\end{aligned}
\tag{B.30}
$$

Difficulties in filling the constraints in equations B.18 and B.30

Relation between elementary filters and global filter is established by equations B.18 and B.30. The terms defined by these equations have to be constant (as normalisation constants) so that multiplication of elementary filters answers to the questions are equivalent to global filter answers.

Most of the terms in these equations are already normalisation constants resulting from asking questions to a joint distribution, but it is interesting to notice that this constant includes $P(M^t|\pi_{lf})$ from global filter and $\prod_{i=1}^{N_i} P(M^{t-1}|\pi_i)$ from elementary filters. In order to keep the necessary conditions required by equations B.18 and B.30, it is recommended to have these terms in all elementary filters equal to uniform distributions.

Appendix C

Behaviour Selection

Behavior selection was proposed in bayes filter in chapter 6: this appendix presents the mathematical development of this filter.

Section C.1 presents the proof of calculation loop for an elementary filter. In section C.2, it is shown that a global filter also presents this calculation loop. In order to show that it is equivalent to use elementary filters and to use global filter in section C.3.

C.1 Elementary Filter

In this section, it is shown that for an elementary filter described by the joint decomposition shown in C.1.1, it is possible to estimate states recursively, using a prediction-estimation calculation loop.

Initially, expression solving the prediction question is found. Then, the expression to answer the behavior selection question is developed and found to be function of prediction expression. Next, estimation expression, also in function of prediction, is developed and the motor question answer is found to be function of estimation expression. Finally, prediciton is shown to be calculated from estimation expression at previous time step.

C.1.1 Joint Distribution

$$
\begin{aligned}
&P(M^{0:t} S_i^{0:t} Z_i^{0:t} B^{0:t} \lambda_i^{0:t} \beta_i^{0:t} | \pi_i) \\
&= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i) \end{array} \right.
\end{aligned}
\tag{C.1}
$$

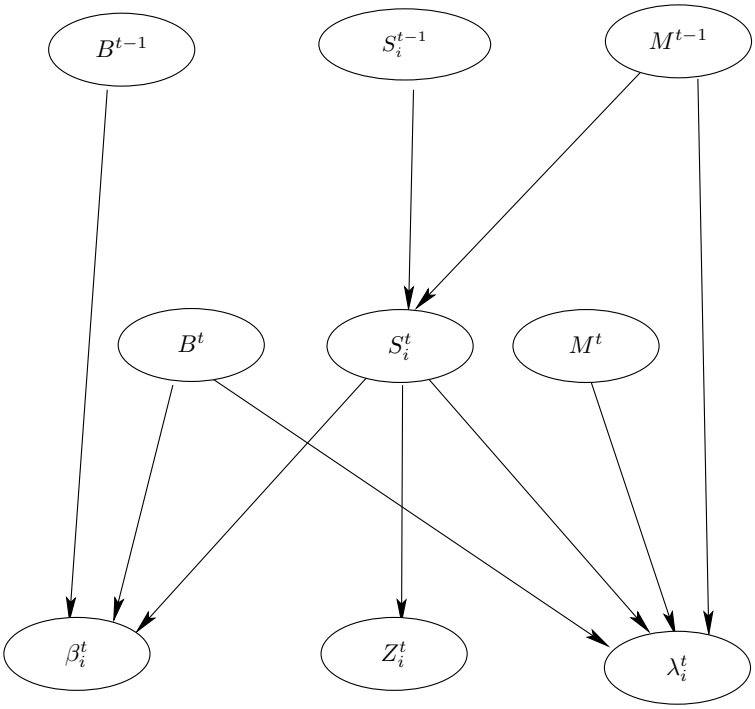


Figure C.1: Bayes Network for Elementary Filter defined in joint distribution equation C.1.

C.1.2 Prediction

Prediction expression is found employing sum and normalization rules, described in appendix A

$$
\begin{aligned}
&P(S_i^t | z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i) \\
&= \left| \begin{array}{l} \frac{1}{P(z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} | \pi_i)} \\ \times \sum \begin{bmatrix} S_i^{0:t-1} \\ Z_i^t \\ B^t \\ M^t \\ \lambda_i^t \\ \beta_i^t \end{bmatrix} \begin{bmatrix} \prod_{j=1}^{t} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{bmatrix} \\ \times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i) \end{bmatrix} \end{array} \right.
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} | \pi_i)} \\
\times \sum \begin{bmatrix} S_i^{0:t-1} \\ Z_i^t \\ B^t \\ M^t \\ \lambda_i^t \\ \beta_i^t \end{bmatrix} \begin{bmatrix}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(Z_i^t | S_i^t \pi_i) \\
\times P(B^t | \pi_i) \times P(\beta_i^t | B^t b^{t-1} S_i^t \pi_i) \\
\times P(M^t | \pi_i) \times P(\lambda_i^t | M^t S_i^t B^t m^{t-1} \pi_i) \\
\times \prod_{j=1}^{t-1} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{bmatrix} \\
\times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i)
\end{bmatrix} \right. \\
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} | \pi_i)} \\
\times \sum_{S_i^{0:t-1}} \begin{bmatrix}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times \sum_{Z_i^t} P(Z_i^t | S_i^t \pi_i) \\
\times \sum_{B^t} \begin{bmatrix} P(B^t | \pi_i) \times \sum_{\beta_i^t} P(\beta_i^t | B^t b^{t-1} S_i^t \pi_i) \\ \times \sum_{M^t} \left[P(M^t | \pi_i) \times \sum_{\lambda_i^t} P(\lambda_i^t | M^t S_i^t B^t m^{t-1} \pi_i) \right] \end{bmatrix} \\
\times \prod_{j=1}^{t-1} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{bmatrix} \\
\times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i)
\end{bmatrix} \right. \\
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} | \pi_i)} \\
\times \sum_{S_i^{0:t-1}} \begin{bmatrix}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times \prod_{j=1}^{t-1} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{bmatrix} \\
\times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i)
\end{bmatrix} \right.
\end{aligned}
\tag{C.2}
$$

C.1.3 Behavior Selection

Behavior selection question is developed using sum and normalization rules, and rearranging terms in order to have it in function of prediction expression.

$$P(B^t | z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \pi_i)$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} | \pi_i)} \\
\times \sum \begin{bmatrix} S_i^{0:t} \\ M^t \\ \lambda_i^t \end{bmatrix} \begin{bmatrix} \prod_{j=1}^{t} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{bmatrix} \\ \times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i) \end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} | \pi_i)} \\
\times \sum \begin{bmatrix} S_i^{0:t} \\ M^t \\ \lambda_i^t \end{bmatrix} \begin{bmatrix} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(z_i^t | S_i^t \pi_i) \\ \times P(B^t | \pi_i) \times P(\beta_i^t | B^t b^{t-1} S_i^t \pi_i) \\ \times P(M^t | \pi_i) \times P(\lambda_i^t | M^t S_i^t B^t m^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{bmatrix} \\ \times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i) \end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} | \pi_i)} \\
\times \sum_{S_i^{0:t}} \begin{bmatrix} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(z_i^t | S_i^t \pi_i) \\ \times P(B^t | \pi_i) \times P(\beta_i^t | B^t b^{t-1} S_i^t \pi_i) \\ \times \sum_{M^t} \left[P(M^t | \pi_i) \times \sum_{\lambda_i^t} P(\lambda_i^t | M^t S_i^t B^t m^{t-1} \pi_i) \right] \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{bmatrix} \\ \times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i) \end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} | \pi_i)} \\
\times P(B^t | \pi_i) \times \\
\sum_{S_i^t} \begin{bmatrix} P(\beta_i^t | B^t b^{t-1} S_i^t \pi_i) \\ \times P(z_i^t | S_i^t \pi_i) \\ \times \sum_{S_i^{0:t-1}} \begin{bmatrix} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{bmatrix} \\ \times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i) \end{bmatrix} \end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l} \frac{P(z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} | \pi_i)}{P(z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} | \pi_i)} \times P(B^t | \pi_i) \\ \times \sum_{S_i^t} \left[\begin{array}{l} P(\beta_i^t | B^t b^{t-1} S_i^t \pi_i) \\ \times P(z_i^t | S_i^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i) \end{array} \right] \end{array} \right. \tag{C.3}
$$

C.1.4 Estimation

Answer to estimation question is found employing sum and normalization rule, and then rearranging terms so that we find estimation in function of prediction expression.

$$
\begin{aligned}
& P(S_i^t | z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \pi_i) \\
&= \left| \begin{array}{l} \frac{1}{P(z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} | \pi_i)} \\ \times \sum \left[\begin{array}{l} S_i^{0:t-1} \\ M^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l} \prod_{j=1}^t \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{1}{P(z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} | \pi_i)} \\ \times \sum \left[\begin{array}{l} S_i^{0:t-1} \\ M^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(z_i^t | S_i^t \pi_i) \\ \times P(B^t | \pi_i) \times P(\beta_i^t | B^t b^{t-1} S_i^t \pi_i) \\ \times P(M^t | \pi_i) \times P(\lambda_i^t | M^t S_i^t B^t m^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{1}{P(z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} | \pi_i)} \\ \times \sum_{S_i^{0:t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(z_i^t | S_i^t \pi_i) \\ \times P(b^t | \pi_i) \times P(\beta_i^t | b^t b^{t-1} S_i^t \pi_i) \\ \times \sum_{M^t} \left[P(M^t | \pi_i) \times \sum_{\lambda_i^t} P(\lambda_i^t | M^t S_i^t b^t m^{t-1} \pi_i) \right] \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i) \end{array} \right] \end{array} \right.
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} | \pi_i)} \\
\times P(z_i^t | S_i^t \pi_i) \\
\times P(b^t | \pi_i) \times P(\beta_i^t | b^t b^{t-1} S_i^t \pi_i) \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i)
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} | \pi_i)}{P(z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} | \pi_i)} \\
\times P(z_i^t | S_i^t \pi_i) \times P(b^t | \pi_i) \times P(\beta_i^t | b^t b^{t-1} S_i^t \pi_i) \\
\times P(S_i^t | z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i)
\end{array} \right|
\end{aligned}
\tag{C.4}
$$

C.1.5 Motor Commands

Applying estimation expression, it is possible to find the final expression for the motor question. Sum and normalization rules were used here.

$$
\begin{aligned}
&P(M^t | z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t} \beta_i^{0:t} \pi_i) \\
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t} \beta_i^{0:t} | \pi_i)} \\
\times \sum_{S_i^{0:t}} \left[\begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i)
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t} \beta_i^{0:t} | \pi_i)} \\
\times \sum_{S_i^{0:t}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(z_i^t | S_i^t \pi_i) \\
\times P(b^t | \pi_i) \times P(\beta_i^t | b^t b^{t-1} S_i^t \pi_i) \\
\times P(M^t | \pi_i) \times P(\lambda_i^t | M^t S_i^t b^t m^{t-1} \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t} \beta_i^{0:t} | \pi_i)} \\
\times P(b^t|\pi_i) \times P(M^t|\pi_i) \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P(\beta_i^t | b^t b^{t-1} S_i^t \pi_i) \\
\times P(\lambda_i^t | M^t S_i^t b^t m^{t-1} \pi_i) \\
\times P(z_i^t | S_i^t \pi_i) \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j \pi_i) \\
\times P(B^j|\pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\
\times P(M^j|\pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i)
\end{array} \right]
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} | \pi_i)}{P(z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t} \beta_i^{0:t} | \pi_i)} \\
\times P(M^t|\pi_i) \times \sum_{S_i^t} \left[\begin{array}{l}
P(\lambda_i^t | M^t S_i^t b^t \pi_i) \\
\times P(S_i^t | z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
\tag{C.5}
$$

C.1.6 Prediction related to Estimation

In order to have a calculation loop, it is desirable to have prediciton as a function of estimation expression of previous time step.

$$
\begin{aligned}
&P(S_i^t | z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i) \\
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} | \pi_i)} \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j \pi_i) \\
\times P(B^j|\pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\
\times P(M^j|\pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i)
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} | \pi_i)} \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(S_i^{t-1} | S_i^{t-2} m^{t-2} \pi_i) \\
\times P(z_i^{t-1} | S_i^{t-1} \pi_i) \\
\times P(b^{t-1}|\pi_i) \times P(\beta_i^{t-1} | b^{t-1} b^{t-2} S_i^{t-1} \pi_i) \\
\times P(m^{t-1}|\pi_i) \times P(\lambda_i^{t-1} | m^{t-1} S_i^{t-1} b^{t-1} m^{t-2} \pi_i) \\
\times \prod_{j=1}^{t-2} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j \pi_i) \\
\times P(B^j|\pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\
\times P(M^j|\pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} | \pi_i)} \\
\times P(b^{t-1}|\pi_i) \times P(m^{t-1}|\pi_i) \\
\times \sum_{S_i^{t-1}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(\beta_i^{t-1} | b^{t-1} b^{t-2} S_i^{t-1} \pi_i) \\
\times P(\lambda_i^{t-1} | m^{t-1} S_i^{t-1} b^{t-1} m^{t-2} \pi_i) \\
\times P(z_i^{t-1} | S_i^{t-1} \pi_i) \\
\times \sum_{S_i^{0:t-2}} \left[\begin{array}{l}
P(S_i^{t-1} | S_i^{t-2} m^{t-2} \pi_i) \\
\times \prod_{j=1}^{t-2} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 Z_i^0 B^0 \lambda_i^0 \beta_i^0 | \pi_i)
\end{array} \right]
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(z_i^{0:t-1} b^{0:t-1} m^{0:t-2} \lambda_i^{0:t-2} \beta_i^{0:t-1} | \pi_i)}{P(z_i^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} | \pi_i)} \\
\times P(m^{t-1}|\pi_i) \\
\times \sum_{S_i^{t-1}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(\lambda_i^{t-1} | m^{t-1} S_i^{t-1} b^{t-1} m^{t-2} \pi_i) \\
\times P(S^{t-1} | z_i^{0:t-1} b^{0:t-1} m^{0:t-2} \lambda_i^{0:t-2} \beta_i^{0:t-1} \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
\tag{C.6}
$$

C.2 Global Filter

This section aims to verify that a global filter described as in section C.2.1 provides a recursive estimation of states.

C.2.1 Joint Distribution

$$
\begin{aligned}
&P(M^{0:t} S^{0:t} Z^{0:t} B^{0:t} \lambda^{0:t} \beta^{0:t} \pi_{bs}) \\
&= \left| \begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\prod_{i=1}^{N_i} P(Z_i^j | S_i^j \pi_i) \\
\times P(B^j | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \right] \\
\times P(M^j | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \right]
\end{array} \right] \\
\times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0 | \pi_{bs})
\end{array} \right|
\end{aligned}
\tag{C.7}
$$

C.2.2 Prediction

Prediction expression is found applying sum and normalization rules.

$$
P(S^t | z^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t-1} \pi_{bs})
$$

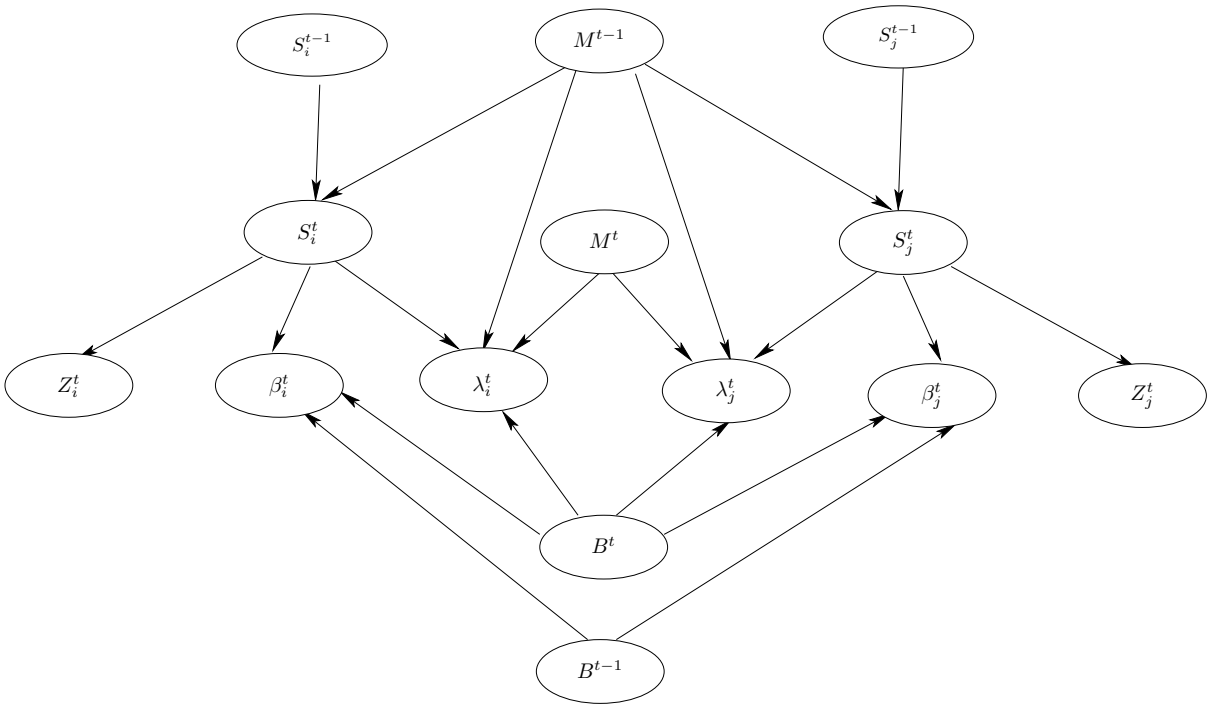


Figure C.2: Bayes Network for Global Filter defined in joint distribution equation C.7.

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t-1} | \pi_{bs})} \\
\times \sum \begin{bmatrix} S^{0:t-1} \\ Z^t \\ B^t \\ M^t \\ \lambda^t \\ \beta^t \end{bmatrix} \left[\begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\prod_{i=1}^{N_i} P(Z_i^j | S_i^j \pi_i) \\
\times P(B^j | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \right] \\
\times P(M^j | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \right]
\end{array} \right] \\
\times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0 | \pi_{bs})
\end{array} \right]
\end{array} \right|
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\
\times \sum \begin{bmatrix} S^{0:t-1} \\ Z^t \\ B^t \\ M^t \\ \lambda^t \\ \beta^t \end{bmatrix} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z_i^t|S_i^t\pi_i) \\
\times P(B^t|\pi_{bs}) \prod_{i=1}^{N_i} [P(\beta_i^t|B^t b^{t-1} S_i^t \pi_i)] \\
\times P(M^t|\pi_{bs}) \prod_{i=1}^{N_i} [P(\lambda_i^t|M^t S_i^t B^t m^{t-1}\pi_i)] \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z_i^j|S_i^j\pi_i) \\
\times P(B^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j|B^j B^{j-1} S_i^j \pi_i)\right] \\
\times P(M^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^j S_i^j B^j M^{j-1}\pi_i)\right]
\end{bmatrix} \\
\times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0|\pi_{bs})
\end{bmatrix}
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\
\times \sum_{S^{0:t-1}} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\times \sum_{Z^t} \left[\prod_{i=1}^{N_i} P(Z_i^t|S_i^t\pi_i)\right] \\
\times \sum_{B^t} \begin{bmatrix}
P(B^t|\pi_{bs}) \times \sum_{\beta_i^t} \prod_{i=1}^{N_i} [P(\beta_i^t|B^t b^{t-1} S_i^t \pi_i)] \\
\times \sum_{M^t} \left[P(M^t|\pi_{bs}) \times \sum_{\lambda_i^t} \prod_{i=1}^{N_i} P(\lambda_i^t|M^t S_i^t B^t m^{t-1}\pi_i)\right]
\end{bmatrix} \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z_i^j|S_i^j\pi_i) \\
\times P(B^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j|B^j B^{j-1} S_i^j \pi_i)\right] \\
\times P(M^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^j S_i^j B^j M^{j-1}\pi_i)\right]
\end{bmatrix} \\
\times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0|\pi_{bs})
\end{bmatrix}
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\
\times \sum_{S^{0:t-1}} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}mt-1\pi_i) \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z_i^j|S_i^j\pi_i) \\
\times P(B^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j|B^j B^{j-1} S_i^j \pi_i)\right] \\
\times P(M^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^j S_i^j B^j M^{j-1}\pi_i)\right]
\end{bmatrix} \\
\times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0|\pi_{bs})
\end{bmatrix}
\end{array} \right|
\end{aligned}
\tag{C.8}
$$

C.2.3 Behavior Selection

Behavior selection expression is developed employing sum and normalozation rules, and rearranging terms in order to have it in function of prediction.

$$P(B^t|z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\pi_{bs})$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t} | \pi_{bs})} \\
\times \sum \begin{bmatrix} S^{0:t} \\ M^t \\ \lambda^t \end{bmatrix} \begin{bmatrix} \prod_{j=1}^{t} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \prod_{i=1}^{N_i} P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \right] \\ \times P(M^j | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \right] \end{bmatrix} \\ \times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0 | \pi_{bs}) \end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t} | \pi_{bs})} \\
\times \sum \begin{bmatrix} S^{0:t} \\ M^t \\ \lambda^t \end{bmatrix} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \prod_{i=1}^{N_i} P(z_i^t | S_i^t \pi_i) \\ \times P(B^t | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^t | B^t b^{t-1} S_i^t \pi_i) \right] \\ \times P(M^t | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^t | M^t S_i^t B^t m^{t-1} \pi_i) \right] \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \prod_{i=1}^{N_i} P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \right] \\ \times P(M^j | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \right] \end{bmatrix} \\ \times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0 | \pi_{bs}) \end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t} | \pi_{bs})} \\
\times \sum_{S^{0:t}} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \prod_{i=1}^{N_i} P(z_i^t | S_i^t \pi_i) \\ \times P(B^t | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^t | B^t b^{t-1} S_i^t \pi_i) \right] \\ \times \sum_{M^t} \left[P(M^t | \pi_{bs}) \times \sum_{\lambda^t} \prod_{i=1}^{N_i} \left[P(\lambda_i^t | M^t S_i^t B^t m^{t-1} \pi_i) \right] \right] \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \prod_{i=1}^{N_i} P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \right] \\ \times P(M^j | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \right] \end{bmatrix} \\ \times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0 | \pi_{bs}) \end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t} | \pi_{bs})} \times P(B^t | \pi_{bs}) \times \\
\sum_{S^t} \begin{bmatrix} \prod_{i=1}^{N_i} \left[P(\beta_i^t | B^t b^{t-1} S_i^t \pi_i) \right] \\ \prod_{i=1}^{N_i} P(z_i^t | S_i^t \pi_i) \\ \times \sum_{S^{0:t-1}} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \prod_{i=1}^{N_i} P(Z_i^j | S_i^j \pi_i) \\ \times P(B^j | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j | B^j B^{j-1} S_i^j \pi_i) \right] \\ \times P(M^j | \pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j | M^j S_i^j B^j M^{j-1} \pi_i) \right] \end{bmatrix} \\ \times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0 | \pi_{bs}) \end{bmatrix} \end{bmatrix}
\end{array} \right|
$$

$$= \left| \begin{array}{l} \frac{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})}{P(z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs})} \times P(B^t|\pi_{bs}) \\ \times \sum_{S^t} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P(\beta_i^t|B^t b^{t-1} S_i^t \pi_i)\right] \\ \prod_{i=1}^{N_i} P(z_i^t|S_i^t\pi_i) \\ \times P(S^t|z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\pi_{bs}) \end{array} \right] \end{array} \right| \tag{C.9}$$

C.2.4 Estimation

Estimation expression is found in function of prediction expression, applying sum rule.

$$P(S^t|z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\pi_{bs})$$

$$= \left| \begin{array}{l} \frac{1}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs})} \\ \times \sum \left[\begin{array}{l} S^{0:t-1} \\ M^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P(Z_i^j|S_i^j\pi_i) \\ \times P(B^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j|B^j B^{j-1} S_i^j \pi_i)\right] \\ \times P(M^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^j S_i^j B^j M^{j-1}\pi_i)\right] \end{array} \right] \\ \times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0|\pi_{bs}) \end{array} \right] \end{array} \right|$$

$$= \left| \begin{array}{l} \frac{1}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs})} \\ \times \sum \left[\begin{array}{l} S^{0:t-1} \\ M^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P(Z_i^t|S_i^t\pi_i) \\ \times P(b^t|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^t|b^t b^{t-1} S_i^t \pi_i)\right] \\ \times P(M^t|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^t|M^t S_i^t B^t m^{t-1}\pi_i)\right] \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P(Z_i^j|S_i^j\pi_i) \\ \times P(B^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j|B^j B^{j-1} S_i^j \pi_i)\right] \\ \times P(M^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^j S_i^j B^j M^{j-1}\pi_i)\right] \end{array} \right] \\ \times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0|\pi_{bs}) \end{array} \right] \end{array} \right|$$

$$= \left| \begin{array}{l} \frac{1}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs})} \\ \times \sum_{S^{0:t-1}} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P(z_i^t|S_i^t\pi_i) \\ \times P(b^t|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^t|b^t b^{t-1} S_i^t \pi_i)\right] \\ \times \sum_{M^t} \left[P(M^t|\pi_{bs}) \times \sum_{\lambda^t} \prod_{i=1}^{N_i} P(\lambda_i^t|M^t S_i^t B^t m^{t-1}\pi_i)\right] \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P(Z_i^j|S_i^j\pi_i) \\ \times P(B^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j|B^j B^{j-1} S_i^j \pi_i)\right] \\ \times P(M^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^j S_i^j B^j M^{j-1}\pi_i)\right] \end{array} \right] \\ \times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0|\pi_{bs}) \end{array} \right] \end{array} \right|$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs})} \prod_{i=1}^{N_i} P(z_i^t|S_i^t\pi_i) \\
\times P(b^t|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^t|b^t b^{t-1} S_i^t \pi_i)\right] \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z_i^j|S_i^j\pi_i) \\
\times P(B^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j|B^j B^{j-1} S_i^j \pi_i)\right] \\
\times P(M^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^j S_i^j B^j M^{j-1}\pi_i)\right]
\end{array} \right] \\
\times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0|\pi_{bs})
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs})} \prod_{i=1}^{N_i} P(z_i^t|S_i^t\pi_i) \\
\times P(b^t|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^t|b^t b^{t-1} S_i^t \pi_i)\right] \\
\times P(S^t|z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\pi_{bs})
\end{array} \right|
\end{aligned}
\tag{C.10}
$$

C.2.5 Motor Commands

Answer to motor command question is an expression in function of estimation expression.

$$
\begin{aligned}
&P(M^t|z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\pi_{bs}) \\
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}|\pi_{bs})} \\
\times \sum_{S^{0:t}} \left[\begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z_i^j|S_i^j\pi_i) \\
\times P(B^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j|B^j B^{j-1} S_i^j \pi_i)\right] \\
\times P(M^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^j S_i^j B^j M^{j-1}\pi_i)\right]
\end{array} \right] \\
\times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0|\pi_{bs})
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}|\pi_{bs})} \\
\times \sum_{S^{0:t}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(z_i^t|S_i^t\pi_i) \\
\times P(b^t|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^t|b^t b^{t-1} S_i^t \pi_i)\right] \\
\times P(M^t|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^t|M^t S_i^t b^t m^{t-1}\pi_i)\right] \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z_i^j|S_i^j\pi_i) \\
\times P(B^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j|B^j B^{j-1} S_i^j \pi_i)\right] \\
\times P(M^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^j S_i^j B^j M^{j-1}\pi_i)\right]
\end{array} \right] \\
\times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0|\pi_{bs})
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}|\pi_{bs})} \times P(b^t|\pi_{bs}) \times P(M^t|\pi_{bs}) \\
\times \sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(\beta_i^t|b^t b^{t-1} S_i^t \pi_i)\right] \\
\prod_{i=1}^{N_i} \left[P(\lambda_i^t|M^t S_i^t b^t m^{t-1} \pi_i)\right] \\
\prod_{i=1}^{N_i} P(z_i^t|S_i^t \pi_i) \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1} m^{t-1} \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1} M^{j-1} \pi_i) \\
\prod_{i=1}^{N_i} P(Z_i^j|S_i^j \pi_i) \\
\times P(B^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j|B^j B^{j-1} S_i^j \pi_i)\right] \\
\times P(M^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^j S_i^j B^j M^{j-1} \pi_i)\right]
\end{array} \right] \\
\times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0|\pi_{bs})
\end{array} \right]
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs})}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}|\pi_{bs})} \times P(M^t|\pi_{bs}) \\
\times \sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(\lambda_i^t|M^t S_i^t b^t m^{t-1} \pi_i)\right] \\
\times P(S^t|z^{0:t} b^{0:t} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t} \pi_{bs})
\end{array} \right]
\end{array} \right|
\end{aligned}
\tag{C.11}
$$

C.2.6 Prediction using Estimation

In order to have recursive calculation, it is necessary to have prediction in function of estimation of previous time step, which is demonstrated here.

$$
\begin{aligned}
&P(S^t|z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\pi_{bs}) \\
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1} m^{t-1} \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1} M^{j-1} \pi_i) \\
\prod_{i=1}^{N_i} P(Z_i^j|S_i^j \pi_i) \\
\times P(B^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j|B^j B^{j-1} S_i^j \pi_i)\right] \\
\times P(M^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^j S_i^j B^j M^{j-1} \pi_i)\right]
\end{array} \right] \\
\times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0|\pi_{bs})
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1} m^{t-1} \pi_i) \\
\prod_{i=1}^{N_i} P(S_i^{t-1}|S_i^{t-2} m^{t-2} \pi_i) \\
\prod_{i=1}^{N_i} P(z_i^{t-1}|S_i^{t-1} \pi_i) \\
\times P(b^{t-1}|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^{t-1}|b^{t-1} b^{t-2} S_i^{t-1} \pi_i)\right] \\
\times P(m^{t-1}|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^{t-1}|m^{t-1} S_i^{t-1} b^{t-1} m^{t-2} \pi_i)\right] \\
\times \prod_{j=1}^{t-2} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1} M^{j-1} \pi_i) \\
\prod_{i=1}^{N_i} P(Z_i^j|S_i^j \pi_i) \\
\times P(B^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j|B^j B^{j-1} S_i^j \pi_i)\right] \\
\times P(M^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^j S_i^j B^j M^{j-1} \pi_i)\right]
\end{array} \right] \\
\times P(M^0 S^0 Z^0 B^0 \lambda^0 \beta^0|\pi_{bs})
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\
\times P(b^{t-1}|\pi_{bs}) \times P(m^{t-1}|\pi_{bs}) \\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} \left[P(\beta_i^{t-1}|b^{t-1}b^{t-2}S_i^{t-1}\pi_i)\right] \\
\prod_{i=1}^{N_i} \left[P(\lambda_i^{t-1}|m^{t-1}S_i^{t-1}b^{t-1}m^{t-2}\pi_i)\right] \\
\prod_{i=1}^{N_i} P(z_i^{t-1}|S_i^{t-1}\pi_i) \\
\times \sum_{S^{0:t-2}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^{t-1}|S_i^{t-2}m^{t-2}\pi_i) \\
\times \prod_{j=1}^{t-2} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z_i^j|S_i^j\pi_i) \\
\times P(B^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\beta_i^j|B^jB^{j-1}S_i^j\pi_i)\right] \\
\times P(M^j|\pi_{bs}) \prod_{i=1}^{N_i} \left[P(\lambda_i^j|M^jS_i^jB^jM^{j-1}\pi_i)\right]
\end{array} \right] \\
\times P(M^0S^0Z^0B^0\lambda^0\beta^0|\pi_{bs})
\end{array} \right]
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(z^{0:t-1}b^{0:t-1}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-1}|\pi_{bs})}{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \times P(m^{t-1}|\pi_{bs}) \\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} \left[P(\lambda_i^{t-1}|m^{t-1}S_i^{t-1}b^{t-1}m^{t-2}\pi_i)\right] \\
\times P(S^{t-1}|z^{0:t-1}b^{0:t-1}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-1}\pi_{bs})
\end{array} \right]
\end{array} \right| \tag{C.12}
$$

C.3 Relations Between Elementary and Global Filter

Here, we aim to show that questions in global filter are proportional to the product of equivalent questions in elementary filter. So we want to show that:

$$
\begin{aligned}
&P(S^t|z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\pi_{bs}) \\
&\quad \propto \prod_{i=1}^{N_i} P(S_i^t|z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\pi_i)
\end{aligned} \tag{C.13}
$$

$$
\begin{aligned}
&P(B^t|z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\pi_{bs}) \\
&\quad \propto \prod_{i=1}^{N_i} P(B^t|z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\pi_i)
\end{aligned} \tag{C.14}
$$

$$
\begin{aligned}
&P(S^t|z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\pi_{bs}) \\
&\quad \propto \prod_{i=1}^{N_i} P(S_i^t|z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\pi_i)
\end{aligned} \tag{C.15}
$$

$$
\begin{aligned}
&P(M^t|z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\pi_{bs}) \\
&\quad \propto \prod_{i=1}^{N_i} P(M^t|z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t}\beta_i^{0:t}\pi_i)
\end{aligned} \tag{C.16}
$$

Lets begin considering that prediction at time instant $t-1$ is proportional to elementary filters prediction:

$$\begin{aligned} & P(S^{t-1}|z^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\pi_{bs}) \\ & = \kappa_{t-2}\prod_{i=1}^{N_i} P(S_i^{t-1}|z_i^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\pi_i) \quad , \end{aligned} \tag{C.17}$$

where κ_{t-2} is a constant calculated from values of $z_i^{0:t-2}$, $b^{0:t-2}$, $m^{0:t-2}$, $\lambda_i^{0:t-2}$ and $\beta_i^{0:t-2}$.

Following calculation loop as stated in equation C.10, we have that estimation at time instant $t-1$ is:

$$\begin{aligned} & P(S^{t-1}|z^{0:t-1}b^{0:t-1}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-1}\pi_{bs}) \\ & = \left| \begin{array}{l} \frac{P(z^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_{bs})}{P(z^{0:t-1}b^{0:t-1}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-1}|\pi_{bs})}\prod_{i=1}^{N_i} P(z_i^{t-1}|S_i^{t-1}\pi_i) \\ \times P(b^{t-1}|\pi_{bs})\prod_{i=1}^{N_i}\left[P(\beta_i^{t-1}|b^{t-1}b^{t-2}S_i^{t-1}\pi_i)\right] \\ \times P(S^{t-1}|z^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\pi_{bs}) \end{array} \right. \\ & = \left| \begin{array}{l} \frac{P(z^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_{bs})}{P(z^{0:t-1}b^{0:t-1}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-1}|\pi_{bs})} \times P(b^{t-1}|\pi_{bs}) \\ \prod_{i=1}^{N_i} P(z_i^{t-1}|S_i^{t-1}\pi_i) \\ \prod_{i=1}^{N_i}\left[P(\beta_i^{t-1}|b^{t-1}b^{t-2}S_i^{t-1}\pi_i)\right] \\ \times P(S^{t-1}|z^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\pi_{bs}) \end{array} \right. \end{aligned} \tag{C.18}$$

And, according to equation C.4, we can conclude that:

$$\begin{aligned} & P(S^{t-1}|z^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\pi_{bs}) \\ & = \left| \begin{array}{l} \frac{P(z^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_{bs})}{P(z^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_{bs})} \\ \prod_{i=1}^{N_i}\left[\frac{P(z_i^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}\right] \times \kappa_{t-2} \\ \prod_{i=1}^{N_i} P(S_i^{t-1}|z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\pi_i) \end{array} \right. \end{aligned} \tag{C.19}$$

Applying equation above to calculate prediction at time instant t as stated in equation C.12, we have:

$$P(S^t|z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\pi_{bs})$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{P(z^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_{bs})}{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})\times P(z^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_{bs})} \\
\times P(b^{t-1}|\pi_{bs}) \times P(m^{t-1}|\pi_{bs}) \\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}mt-1\pi_i) \\
\prod_{i=1}^{N_i} \left[P(\beta_i^{t-1}|b^{t-1}b^{t-2}S_i^{t-1}\pi_i)\right] \\
\prod_{i=1}^{N_i} \left[P(\lambda_i^{t-1}|m^{t-1}S_i^{t-1}b^{t-1}\pi_i)\right] \\
\times \frac{P(z^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_{bs})}{P(z^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_{bs})} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}\right] \times \kappa_{t-2} \\
\prod_{i=1}^{N_i} P(S_i^{t-1}|z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\pi_i)
\end{array} \right]
\end{array} \right. \\
&= \left| \begin{array}{l}
\frac{\kappa_{t-2}}{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}\right] \\
\times P(b^{t-1}|\pi_{bs}) \times P(m^{t-1}|\pi_{bs}) \times \\
\sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}mt-1\pi_i) \\
\prod_{i=1}^{N_i} \left[P(\beta_i^{t-1}|b^{t-1}b^{t-2}S_i^{t-1}\pi_i)\right] \\
\prod_{i=1}^{N_i} \left[P(\lambda_i^{t-1}|m^{t-1}S_i^{t-1}b^{t-1}\pi_i)\right] \\
\prod_{i=1}^{N_i} P(S_i^{t-1}|z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\pi_i)
\end{array} \right]
\end{array} \right. \\
&= \left| \begin{array}{l}
\frac{\kappa_{t-2}}{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}\right] \\
\times P(b^{t-1}|\pi_{bs}) \times P(m^{t-1}|\pi_{bs}) \\
\times \sum_{S^{t-1}} \left[\prod_{i=1}^{N_i} \left\{ \begin{array}{l}
P(S_i^t|S_i^{t-1}mt-1\pi_i) \\
\times P(\beta_i^{t-1}|b^{t-1}b^{t-2}S_i^{t-1}\pi_i) \\
\times P(\lambda_i^{t-1}|m^{t-1}S_i^{t-1}b^{t-1}\pi_i) \\
\times P(S_i^{t-1}|z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\pi_i)
\end{array} \right\} \right]
\end{array} \right.
\end{aligned}
\tag{C.20}
$$

From equation C.6, we can see that prediction at time t is proportional to the product of prediction in each elementary filter:

$$
\begin{aligned}
&P(S^t|z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\pi_{bs}) \\
&= \left| \begin{array}{l}
\frac{\kappa_{t-2}}{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}\right] \\
\times \frac{P(b^{t-1}|\pi_{bs})\times P(m^{t-1}|\pi_{bs})}{\prod_{i=1}^{N_i}[P(b^{t-1}|\pi_i)\times P(m^{t-1}|\pi_i)]} \\
\prod_{i=1}^{N_i} \left[\sum_{S^{t-1}} \left\{ \begin{array}{l}
P(S_i^t|S_i^{t-1}mt-1\pi_i) \\
\times P(b^{t-1}|\pi_i) \times P(\beta_i^{t-1}|b^{t-1}b^{t-2}S_i^{t-1}\pi_i) \\
\times P(m^{t-1}|\pi_i) \times P(\lambda_i^{t-1}|m^{t-1}S_i^{t-1}b^{t-1}\pi_i) \\
\times P(S_i^{t-1}|z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\pi_i)
\end{array} \right\} \right]
\end{array} \right.
\end{aligned}
\tag{C.21}
$$

$$= \left| \begin{array}{l} \frac{\kappa_{t-2}}{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\ \prod_{i=1}^{N_i}\left[\frac{P(z_i^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}\right] \\ \prod_{i=1}^{N_i}\left[\frac{P(z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}|\pi_i)\times P(z_i^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}|\pi_i)}\right] \\ \times\frac{P(b^{t-1}|\pi_{bs})\times P(m^{t-1}|\pi_{bs})}{\prod_{i=1}^{N_i}[P(b^{t-1}|\pi_i)\times P(m^{t-1}|\pi_i)]} \\ \prod_{i=1}^{N_i}P(S_i^t|z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\pi_i) \end{array} \right. \tag{C.22}$$

From equation above, we can also know how to calculate κ_{t-1}:

$$\kappa_{t-1} = \\ = \left| \begin{array}{l} \frac{\kappa_{t-2}}{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\ \prod_{i=1}^{N_i}\left[\frac{P(z_i^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}|\pi_i)}\right] \\ \times\frac{P(b^{t-1}|\pi_{bs})\times P(m^{t-1}|\pi_{bs})}{\prod_{i=1}^{N_i}[P(b^{t-1}|\pi_i)\times P(m^{t-1}|\pi_i)]} \\ \prod_{i=1}^{N_i}\left[\frac{P(z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}|\pi_i)\times P(z_i^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}|\pi_i)}\right] \end{array} \right. \tag{C.23}$$

Now, it is sufficient to show that:

$$P(S^1|z^0b^0m^0\lambda^0\beta^0\pi_{bs}) \propto \prod_{i=1}^{N_i} P(S_i^1|z_i^0b^0m^0\lambda_i^0\beta_i^0\pi_i).$$

Concerning the prediction for an elementary filter at time step $t = 1$ we find the following expression:

$$P(S_i^1|z_i^0b^0m^0\lambda_i^0\beta_i^0\pi_i) \\ = \left| \begin{array}{l} \frac{1}{P(z_i^0b^0m^0\lambda_i^0\beta_i^0|\pi_i)} \\ \times\sum\left[\begin{array}{l} S_i^0 \\ Z_i^1 \\ B^1 \\ M^1 \\ \lambda_i^1 \\ \beta_i^1 \end{array}\right]\left[\begin{array}{l} P(m^0|\pi_i)\times P(S_i^0|\pi_i)\times P(z_i^0|S_i^0\pi_i)\times P(b^0|\pi_i) \\ \times P(\lambda_i^0|m^0S_i^0\pi_i)\times P(\beta_i^0|b^0S_i^0\pi_i) \\ \times P(S_i^1|S_i^0m^0\pi_i) \\ \times P(Z_i^1|S_i^1\pi_i) \\ \times P(B^1|\pi_i)\times P(\beta_i^1|B^1b^0S_i^1\pi_i) \\ \times P(M^1|\pi_i)\times P(\lambda_i^1|m^1S_i^1B^1\pi_i) \end{array}\right] \end{array} \right. \\ = \left| \begin{array}{l} \frac{P(m^0|\pi_i)\times P(b^0|\pi_i)}{P(z_i^0b^0m^0\lambda_i^0\beta_i^0|\pi_i)} \\ \times\sum_{S_i^0}\left[\begin{array}{l} P(S_i^0|\pi_i)\times P(z_i^0|S_i^0\pi_i) \\ \times P(\lambda_i^0|m^0S_i^0\pi_i)\times P(\beta_i^0|b^0S_i^0\pi_i) \\ \times P(S_i^1|S_i^0m^0\pi_i) \end{array}\right] \end{array} \right. \tag{C.24}$$

And we have that the global prediction expression for $t = 1$ is:

$$
\begin{aligned}
&P(S^1|z^0b^0m^0\lambda^0\beta^0\pi_{bs}) \\
&= \left| \begin{array}{l} \frac{1}{P(z^0b^0m^0\lambda^0\beta^0|\pi_{bs})} \\ \times \sum \begin{bmatrix} S^0 \\ Z^1 \\ B^1 \\ M^1 \\ \lambda^1 \\ \beta^1 \end{bmatrix} \left\{ \begin{array}{l} P(m^0|\pi_{bs}) \times P(S^0|\pi_{bs}) \times P(z^0|S^0\pi_{bs}) \times P(b^0|\pi_{bs}) \\ \times P(\lambda^0|m^0S^0\pi_{lf}) \times P(\beta^0|b^0S^0\pi_{bs}) \\ \prod_{i=1}^{N_i} P(S_i^1|S_i^0m^0\pi_i) \\ \prod_{i=1}^{N_i} P(Z_i^1|S_i^1\pi_i) \\ \times P(B^1|\pi_{bs}) \prod_{i=1}^{N_i} [P(\beta_i^1|B^1b^0S_i^1\pi_i)] \\ \times P(M^1|\pi_{bs}) \prod_{i=1}^{N_i} [P(\lambda_i^1|M^1S_i^1B^1\pi_i)] \end{array} \right\} \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(m^0|\pi_{bs}) \times P(b^0|\pi_{bs})}{P(z^0b^0m^0\lambda^0\beta^0|\pi_{bs})} \\ \times \sum_{S^0} \left\{ \begin{array}{l} P(S^0|\pi_{bs}) \times P(z^0|S^0\pi_{bs}) \\ \times P(\lambda^0|m^0S^0\pi_{lf}) \times P(\beta^0|b^0S^0\pi_{bs}) \\ \prod_{i=1}^{N_i} P(S_i^1|S_i^0m^0\pi_i) \end{array} \right\} \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(m^0|\pi_{bs}) \times P(b^0|\pi_{bs})}{P(z^0b^0m^0\lambda^0\beta^0|\pi_{bs})} \\ \times \sum_{S^0} \prod_{i=1}^{N_i} \left\{ \begin{array}{l} P(S_i^0|\pi_i) \times P(z_i^0|S_i^0\pi_i) \\ \times P(\lambda_i^0|m^0S_i^0\pi_i) \times P(\beta_i^0|b^0S_i^0\pi_i) \\ \times P(S_i^1|S_i^0m^0\pi_i) \end{array} \right\} \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(m^0|\pi_{bs}) \times P(b^0|\pi_{bs})}{P(z^0b^0m^0\lambda^0\beta^0|\pi_{bs})} \\ \prod_{i=1}^{N_i} \sum_{S_i^0} \left\{ \begin{array}{l} P(S_i^0|\pi_i) \times P(z_i^0|S_i^0\pi_i) \\ \times P(\lambda_i^0|m^0S_i^0\pi_i) \times P(\beta_i^0|b^0S_i^0\pi_i) \\ \times P(S_i^1|S_i^0m^0\pi_i) \end{array} \right\} \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(m^0|\pi_{bs}) \times P(b^0|\pi_{bs})}{P(z^0b^0m^0\lambda^0\beta^0|\pi_{bs})} \prod_{i=1}^{N_i} \frac{P(z_i^0b^0m^0\lambda_i^0\beta_i^0|\pi_i)}{P(m^0|\pi_i) \times P(b^0|\pi_i)} \\ \prod_{i=1}^{N_i} P(S_i^1|z_i^0b^0m^0\lambda_i^0\beta_i^0\pi_i) \end{array} \right.
\end{aligned}
\tag{C.25}
$$

And it follows that κ_0 can be calculated by:

$$
\kappa_0 = \left| \frac{P(m^0|\pi_{bs}) \times P(b^0|\pi_{bs})}{P(z^0b^0m^0\lambda^0\beta^0|\pi_{bs})} \prod_{i=1}^{N_i} \frac{P(z_i^0b^0m^0\lambda_i^0\beta_i^0|\pi_i)}{P(m^0|\pi_i) \times P(b^0|\pi_i)} \right. \tag{C.26}
$$

Successive values for κ are calculated by equation C.23.
From equation C.9, we can find proportionality for behavior question:

$$
\begin{aligned}
&P(B^t|z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\pi_{bs}) \\
&= \left| \begin{array}{l} \frac{P(z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})}{P(z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs}) \times P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\ \times P(B^t|\pi_{bs}) \times \\ \sum_{S^t} \left[\begin{array}{l} \prod_{i=1}^{N_i} [P(\beta_i^t|B^tb^{t-1}S_i^t\pi_i)] \\ \times P(S^t|z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\pi_{bs}) \end{array} \right] \end{array} \right.
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{P(z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})}{P(z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs})\times P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\
\times P(B^t|\pi_{bs})\times \\
\sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(\beta_i^t|B^t b^{t-1} S_i^t \pi_i)\right] \\
\times \frac{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})}{P(z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)}{P(z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)}\right] \times \kappa_{t-1} \\
\prod_{i=1}^{N_i} P(S_i^t|z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\pi_i)
\end{array} \right]
\end{array} \right. \\
&= \left| \begin{array}{l}
\frac{\kappa_{t-1}\times P(B^t|\pi_{bs})}{P(z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs})} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)}{P(z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)}\right] \\
\prod_{i=1}^{N_i} \sum_{S_i^t} \left[\begin{array}{l}
P(\beta_i^t|B^t b^{t-1} S_i^t \pi_i) \\
\times P(S_i^t|z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\pi_i)
\end{array} \right]
\end{array} \right. \\
&= \left| \begin{array}{l}
\frac{\kappa_{t-1}}{P(z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs})} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)}{P(z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)}\right] \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}|\pi_i)\times P(z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}|\pi_i)}{P(z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}|\pi_i)}\right] \\
\times \frac{P(B^t|\pi_{bs})}{\prod_{i=1}^{N_i}\left[P(B^t|\pi_i)\right]} \prod_{i=1}^{N_i} P(B^t|z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\pi_i)
\end{array} \right.
\end{aligned}
\tag{C.27}
$$

Regarding motor commands, we can show that proportionality also holds, as in the following equations. Global expression for motor commands as shown in equation C.11:

$$
\begin{aligned}
&P(M^t|z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\pi_{bs}) \\
&= \left| \begin{array}{l}
\frac{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs})}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}|\pi_{bs})} \\
\times P(b^t|\pi_{bs}) \times P(M^t|\pi_{bs}) \\
\times \sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(\beta_i^t|b^t b^{t-1} S_i^t \pi_i)\right] \\
\prod_{i=1}^{N_i} \left[P(\lambda_i^t|M^t S_i^t b^t \pi_i)\right] \\
\times P(S^t|z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\pi_{bs})
\end{array} \right]
\end{array} \right. \\
&= \left| \begin{array}{l}
\frac{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs})}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}|\pi_{bs})} \\
\times P(b^t|\pi_{bs}) \times P(M^t|\pi_{bs}) \\
\times \sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P(\beta_i^t|b^t b^{t-1} S_i^t \pi_i)\right] \\
\prod_{i=1}^{N_i} \left[P(\lambda_i^t|M^t S_i^t b^t \pi_i)\right] \\
\times \frac{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})}{P(z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)}{P(z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)}\right] \times \kappa_{t-1} \\
\prod_{i=1}^{N_i} P(S_i^t|z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\pi_i)
\end{array} \right]
\end{array} \right.
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}|\pi_{bs})}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}|\pi_{bs})} \\
\times \frac{P(z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})}{P(z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_{bs})} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)}{P(z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)} \right] \\
\times \kappa_{t-1} \times P(b^t|\pi_{bs}) \times P(M^t|\pi_{bs}) \times \\
\sum_{S^t} \prod_{i=1}^{N_i} \left[\begin{array}{l}
P(\beta_i^t|b^t b^{t-1} S_i^t \pi_i) \\
\times P(\lambda_i^t|M^t S_i^t b^t \pi_i) \\
\times P(S_i^t|z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i)
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{\kappa_{t-1}}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}|\pi_{bs})} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)}{P(z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)} \right] \\
\times \frac{P(b^t|\pi_{bs}) \times P(M^t|\pi_{bs})}{\prod_{i=1}^{N_i} P(b^t|\pi_i) \times P(M^t|\pi_i)} \\
\prod_{i=1}^{N_i} \sum_{S_i^t} \left[\begin{array}{l}
P(b^t|\pi_i) \times P(\beta_i^t|b^t b^{t-1} S_i^t \pi_i) \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t S_i^t b^t \pi_i) \\
\times P(S_i^t|z_i^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \pi_i)
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{\kappa_{t-1}}{P(z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}|\pi_{bs})} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)}{P(z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}|\pi_i)} \right] \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t}\beta_i^{0:t}|\pi_i) \times P(z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}|\pi_i)}{P(z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}|\pi_i)} \right] \\
\times \frac{P(b^t|\pi_{bs}) \times P(M^t|\pi_{bs})}{\prod_{i=1}^{N_i} P(b^t|\pi_i) \times P(M^t|\pi_i)} \\
\prod_{i=1}^{N_i} P(M^t|z_i^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t} \beta_i^{0:t} \pi_i)
\end{array} \right|
\end{aligned}
\tag{C.28}
$$

Appendix D

Attention Selection

D.1 Elementary Filter

D.1.1 Joint Distribution

$$
\begin{aligned}
&P(M^{0:t}S_i^{0:t}Z_i^{0:t}B^{0:t}C^{0:t}\lambda_i^{0:t}\beta_i^{0:t}\alpha_i^{0:t}|\pi_i) \\
&= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\ \times P(Z_i^j|S_i^jC^j\pi_i) \\ \times P(B^j|\pi_i) \times P(\beta_i^j|B^jS_i^jB^{j-1}\pi_i) \\ \times P(C^j|\pi_i) \times P(\alpha_i^j|C^jS_i^jB^j\pi_i) \\ \times P(M^j|\pi_i) \times P(\lambda_i^j|M^jB^jS_i^jM^{j-1}\pi_i) \end{array} \right] \\ \times P(M^0S_i^0C^0B^0Z_i^0\lambda_i^0\beta_i^0\alpha_i^0|\pi_i) \end{array} \right.
\end{aligned}
\tag{D.1}
$$

D.1.2 Prediction

$$
\begin{aligned}
&P(S_i^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}\pi_i) \\
&= \left| \begin{array}{l} \frac{1}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)} \\ \times \sum \left[\begin{array}{l} S_i^{0:t-1} \\ Z_i^t \\ C^t \\ B^t \\ M^t \\ \lambda_i^t \\ \beta_i^t \\ \alpha_i^t \end{array} \right] \left[\begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\ \times P(Z_i^j|S_i^jC^j\pi_i) \\ \times P(B^j|\pi_i) \times P(\beta_i^j|B^jS_i^jB^{j-1}\pi_i) \\ \times P(C^j|\pi_i) \times P(\alpha_i^j|C^jS_i^jB^j\pi_i) \\ \times P(M^j|\pi_i) \times P(\lambda_i^j|M^jB^jS_i^jM^{j-1}\pi_i) \end{array} \right] \\ \times P(M^0S_i^0C^0B^0Z_i^0\lambda_i^0\beta_i^0\alpha_i^0|\pi_i) \end{array} \right] \end{array} \right.
\end{aligned}
$$

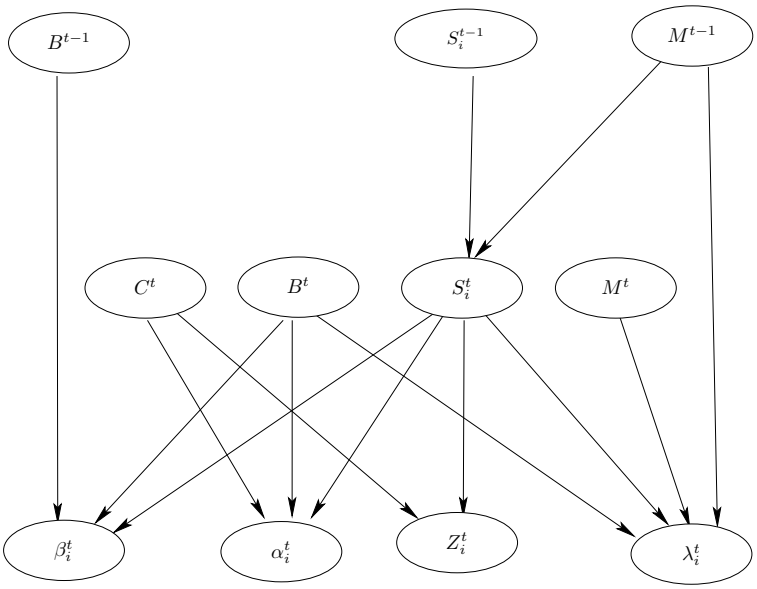


Figure D.1: Bayes Network for Elementary Filter defined in joint distribution D.1.

$$
= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} | \pi_i)} \\
\times \sum \left[\begin{array}{l} S_i^{0:t-1} \\ Z_i^t \\ C^t \\ B^t \\ M^t \\ \lambda_i^t \\ \beta_i^t \\ \alpha_i^t \end{array} \right]
\left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(Z_i^t | S_i^t C^t \pi_i) \\
\times P(B^t | \pi_i) \times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\
\times P(C^t | \pi_i) \times P(\alpha_i^t | C^t S_i^t B^t \pi_i) \\
\times P(M^t | \pi_i) \times P(\lambda_i^t | M^t B^t S_i^t \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} | \pi_i)} \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times \sum_{B^t} \left[\begin{array}{l}
P(B^t|\pi_i) \times \sum_{\beta_i^t} P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\
\times \sum_{C^t} \left[\begin{array}{l} P(C^t|\pi_i) \times \sum_{\alpha_i^t} P(\alpha_i^t | C^t S_i^t B^t \pi_i) \\ \times \sum_{Z_i^t} P(Z^t | S_i^t C^t \pi_i) \end{array} \right] \\
\times \sum_{M^t} \left[P(M^t|\pi_i) \times \sum_{\lambda_i^t} P(\lambda_i^t | M^t B^t S_i^t \pi_i) \right]
\end{array} \right] \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j|\pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j|\pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} | \pi_i)} \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j|\pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j|\pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
\tag{D.2}
$$

D.1.3 Attention Selection

$$
\begin{aligned}
&P(C^t B^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} \pi_i) \\
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\
\times \sum \left[\begin{array}{l} S_i^{0:t} \\ Z_i^t \\ M^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j|\pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j|\pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\
\times \sum \begin{bmatrix} S_i^{0:t} \\ Z_i^t \\ M^t \\ \lambda_i^t \end{bmatrix} \begin{bmatrix}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(Z_i^t | S_i^t C^t \pi_i) \\
\times P(B^t | \pi_i) \times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\
\times P(C^t | \pi_i) \times P(\alpha_i^t | C^t S_i^t B^t \pi_i) \\
\times P(M^t | \pi_i) \times P(\lambda_i^t | M^t B^t S_i^t \pi_i) \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{bmatrix} \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\
\times \sum_{S_i^{0:t}} \begin{bmatrix}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times \sum_{Z_i^t} P(Z_i^t | S_i^t C^t \pi_i) \\
\times P(B^t | \pi_i) \times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\
\times P(C^t | \pi_i) \times P(\alpha_i^t | C^t S_i^t B^t \pi_i) \\
\times \sum_{M^t} \left[P(M^t | \pi_i) \times \sum_{\lambda_i^t} P(\lambda_i^t | M^t B^t S_i^t \pi_i) \right] \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{bmatrix} \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t} | \pi_i)} \\
\times \sum_{S_i^{0:t}} \begin{bmatrix}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times \sum_{Z_i^t} P(Z_i^t | S_i^t C^t \pi_i) \\
\times P(B^t | \pi_i) \times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\
\times P(C^t | \pi_i) \times P(\alpha_i^t | C^t S_i^t B^t \pi_i) \\
\times \sum_{M^t} \left[P(M^t | \pi_i) \times \sum_{\lambda_i^t} P(\lambda_i^t | M^t B^t S_i^t \pi_i) \right] \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{bmatrix} \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{bmatrix}
\end{array} \right|
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P(B^t|\pi_i) \times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\
\times P(C^t|\pi_i) \times P(\alpha_i^t | C^t S_i^t B^t \pi_i) \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j|\pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j|\pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{array} \right]
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} | \pi_i)}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \times P(B^t|\pi_i) \times P(C^t|\pi_i) \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\
\times P(\alpha_i^t | C^t S_i^t B^t \pi_i) \\
\times P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
\tag{D.3}
$$

D.1.4 Behavior Selection

$$
\begin{aligned}
&P(B^t | z_i^{0:t} c^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} \pi_i) \\
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} c^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\
\times \sum_{\substack{S_i^{0:t} \\ M^t \\ \lambda_i^t}} \left[\begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j|\pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j|\pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} c^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\
\times \sum \left[\begin{array}{c} S_i^{0:t} \\ M^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(z_i^t | S_i^t c^t \pi_i) \\
\times P(B^t|\pi_i) \times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\
\times P(c^t|\pi_i) \times P(\alpha_i^t | c^t S_i^t B^t \pi_i) \\
\times P(M^t|\pi_i) \times P(\lambda_i^t | M^t B^t S_i^t \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j|\pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j|\pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} c^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\
\times \sum_{S_i^{0:t}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(z_i^t | S_i^t c^t \pi_i) \\
\times P(B^t | \pi_i) \times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\
\times P(c^t | \pi_i) \times P(\alpha_i^t | c^t S_i^t B^t \pi_i) \\
\times \sum_{M^t} \left[P(M^t | \pi_i) \times \sum_{\lambda_i^t} P(\lambda_i^t | M^t B^t S_i^t \pi_i) \right] \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} c^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \times P(B^t | \pi_i) \times P(c^t | \pi_i) \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P(z_i^t | S_i^t c^t \pi_i) \\
\times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\
\times P(\alpha_i^t | c^t S_i^t B^t \pi_i) \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{array} \right]
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} | \pi_i)}{P(z_i^{0:t} c^{0:t} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \times P(B^t | \pi_i) \times P(c^t | \pi_i) \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P(z_i^t | S_i^t c^t \pi_i) \\
\times P(\beta_i^t | B^t S_i^t b^{t-1} \pi_i) \\
\times P(\alpha_i^t | c^t S_i^t B^t \pi_i) \\
\times P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
\tag{D.4}
$$

D.1.5 Estimation

$$
\begin{aligned}
&P(S_i^t | z_i^{0:t} c^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} \pi_i) \\
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} c^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\
\times \sum \left[\begin{array}{l} S_i^{0:t-1} \\ M^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} c^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\
\times \sum \begin{bmatrix} S_i^{0:t-1} \\ M^t \\ \lambda_i^t \end{bmatrix} \begin{bmatrix}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(z_i^t | S_i^t c^t \pi_i) \\
\times P(b^t | \pi_i) \times P(\beta_i^t | b^t S_i^t b^{t-1} \pi_i) \\
\times P(c^t | \pi_i) \times P(\alpha_i^t | c^t S_i^t b^t \pi_i) \\
\times P(M^t | \pi_i) \times P(\lambda_i^t | M^t b^t S_i^t \pi_i) \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{bmatrix} \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} c^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\
\times \sum_{S_i^{0:t-1}} \begin{bmatrix}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(z_i^t | S_i^t c^t \pi_i) \\
\times P(b^t | \pi_i) \times P(\beta_i^t | b^t S_i^t b^{t-1} \pi_i) \\
\times P(c^t | \pi_i) \times P(\alpha_i^t | c^t S_i^t b^t \pi_i) \\
\times \sum_{M^t} \left[P(M^t | \pi_i) \times \sum_{\lambda_i^t} P(\lambda_i^t | M^t b^t S_i^t \pi_i) \right] \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{bmatrix} \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t} c^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\
\times P(z_i^t | S_i^t c^t \pi_i) \\
\times P(b^t | \pi_i) \times P(\beta_i^t | b^t S_i^t b^{t-1} \pi_i) \\
\times P(c^t | \pi_i) \times P(\alpha_i^t | c^t S_i^t b^t \pi_i) \\
\times \sum_{S_i^{0:t-1}} \begin{bmatrix}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{bmatrix} \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{bmatrix}
\end{array} \right|
$$

$$= \left| \begin{array}{l} \frac{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} | \pi_i)}{P(z_i^{0:t} c^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\ \times P(z_i^t | S_i^t c^t \pi_i) \\ \times P(b^t | \pi_i) \times P(\beta_i^t | b^t S_i^t b^{t-1} \pi_i) \\ \times P(c^t | \pi_i) \times P(\alpha_i^t | c^t S_i^t b^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \end{array} \right. \tag{D.5}$$

D.1.6 Motor Command

$$P(M^t | z_i^{0:t} c^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t} \beta_i^{0:t} \alpha_i^{0:t} \pi_i)$$

$$= \left| \begin{array}{l} \frac{1}{P(z_i^{0:t} c^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\ \times \sum_{S_i^{0:t}} \left[\begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j C^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\ \times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i) \end{array} \right] \end{array} \right.$$

$$= \left| \begin{array}{l} \frac{1}{P(z_i^{0:t} c^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\ \times \sum_{S_i^{0:t}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(z_i^t | S_i^t c^t \pi_i) \\ \times P(b^t | \pi_i) \times P(\beta_i^t | b^t S_i^t b^{t-1} \pi_i) \\ \times P(c^t | \pi_i) \times P(\alpha_i^t | c^t S_i^t b^t \pi_i) \\ \times P(M^t | \pi_i) \times P(\lambda_i^t | M^t b^t S_i^t \pi_i) \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j C^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\ \times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i) \end{array} \right] \end{array} \right.$$

$$= \left| \begin{array}{l} \frac{1}{P(z_i^{0:t} c^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\ \times \sum_{S_i^{t}} \left[\begin{array}{l} P(z_i^t | S_i^t c^t \pi_i) \\ \times P(b^t | \pi_i) \times P(\beta_i^t | b^t S_i^t b^{t-1} \pi_i) \\ \times P(c^t | \pi_i) \times P(\alpha_i^t | c^t S_i^t b^t \pi_i) \\ \times P(M^t | \pi_i) \times P(\lambda_i^t | M^t b^t S_i^t \pi_i) \\ \times \sum_{S_i^{0:t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j C^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\ \times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i) \end{array} \right] \end{array} \right] \end{array} \right.$$

$$= \left| \begin{array}{l} \frac{P(z_i^{0:t} c^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)}{P(z_i^{0:t} c^{0:t} b^{0:t} m^{0:t-1} \lambda_i^{0:t} \beta_i^{0:t} \alpha_i^{0:t} | \pi_i)} \\ \times \sum_{S_i^t} \left[\begin{array}{l} P(M^t|\pi_i) \times P(\lambda_i^t | M^t b^t S_i^t \pi_i) \\ \times P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \end{array} \right] \end{array} \right. \tag{D.6}$$

D.1.7 Prediction using estimation

$$P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i)$$

$$= \left| \begin{array}{l} \frac{1}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} | \pi_i)} \\ \times \sum_{S_i^{0:t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j C^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\ \times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i) \end{array} \right] \end{array} \right.$$

$$= \left| \begin{array}{l} \frac{1}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} | \pi_i)} \\ \times \sum_{S_i^{0:t-1}} \left[\begin{array}{l} P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\ \times P(S_i^{t-1} | S_i^{t-2} m^{t-2} \pi_i) \\ \times P(z_i^{t-1} | S_i^{t-1} c^{t-1} \pi_i) \\ \times P(b^{t-1} | \pi_i) \times P(\beta_i^{t-1} | b^{t-1} S_i^{t-1} b^{t-2} \pi_i) \\ \times P(c^{t-1} | \pi_i) \times P(\alpha_i^{t-1} | c^{t-1} S_i^{t-1} b^{t-1} \pi_i) \\ \times P(m^{t-1} | \pi_i) \times P(\lambda_i^{t-1} | m^{t-1} b^{t-1} S_i^{t-1} \pi_i) \\ \times \prod_{j=1}^{t-2} \left[\begin{array}{l} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\ \times P(Z_i^j | S_i^j C^j \pi_i) \\ \times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\ \times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i) \end{array} \right] \end{array} \right.$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} | \pi_i)} \\
\times P(b^{t-1}|\pi_i) \times P(c^{t-1}|\pi_i) \times P(m^{t-1}|\pi_i) \\
\times \sum_{S_i^{t-1}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(z_i^{t-1} | S_i^{t-1} c^{t-1} \pi_i) \\
\times P(\beta_i^{t-1} | b^{t-1} S_i^{t-1} b^{t-2} \pi_i) \\
\times P(\alpha_i^{t-1} | c^{t-1} S_i^{t-1} b^{t-1} \pi_i) \\
\times P(\lambda_i^{t-1} | m^{t-1} b^{t-1} S_i^{t-1} \pi_i) \\
\times \sum_{S_i^{0:t-2}} \left[\begin{array}{l}
P(S_i^{t-1} | S_i^{t-2} m^{t-2} \pi_i) \\
\times \prod_{j=1}^{t-2} \left[\begin{array}{l}
P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\times P(Z_i^j | S_i^j C^j \pi_i) \\
\times P(B^j | \pi_i) \times P(\beta_i^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j | \pi_i) \times P(\alpha_i^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j | \pi_i) \times P(\lambda_i^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)
\end{array} \right]
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-2} \lambda_i^{0:t-2} \beta_i^{0:t-1} \alpha_i^{0:t-1} | \pi_i)}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} | \pi_i)} \times P(m^{t-1}|\pi_i) \\
\times \sum_{S_i^{t-1}} \left[\begin{array}{l}
P(S_i^t | S_i^{t-1} m^{t-1} \pi_i) \\
\times P(\lambda_i^{t-1} | m^{t-1} b^{t-1} S_i^{t-1} \pi_i) \\
\times P(S_i^{t-1} | z_i^{0:t-1} c^{0:t-1} b^{0:t-2-1} m^{0:t-2} \lambda_i^{0:t-2} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
\tag{D.7}
$$

D.2 Global Filter

D.2.1 Joint Distribution

$$
\begin{aligned}
&P(M^{0:t} S^{0:t} Z^{0:t} B^{0:t} C^{0:t} \lambda^{0:t} \beta^{0:t} \alpha^{0:t} | \pi_{as}) \\
&= \left| \begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j | S_i^{j-1} M^{j-1} \pi_i) \\
\prod_{i=1}^{N_i} P(Z^j | S_i^j C^j \pi_i) \\
\times P(B^j | \pi_{as}) \prod_{i=1}^{N_i} P(\beta^j | B^j S_i^j B^{j-1} \pi_i) \\
\times P(C^j | \pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j | C^j S_i^j B^j \pi_i) \\
\times P(M^j | \pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j | M^j B^j S_i^j M^{j-1} \pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_{as})
\end{array} \right|
\end{aligned}
\tag{D.8}
$$

D.2.2 Prediction

$$
P(S^t | z^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t-1} \alpha^{0:t-1} \pi_{as})
$$

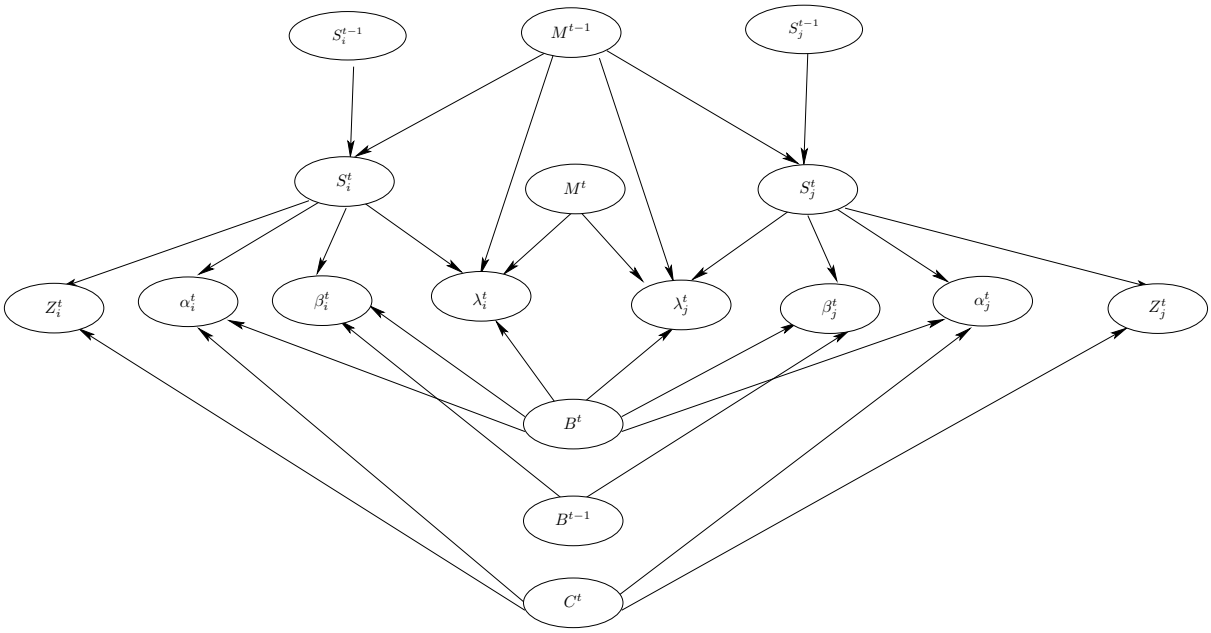


Figure D.2: Bayes Network for Global filter defined in joint distribution D.8.

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \sum \begin{bmatrix} S^{0:t-1} \\ Z^t \\ C^t \\ B^t \\ M^t \\ \lambda^t \\ \beta^t \\ \alpha^t \end{bmatrix} \begin{bmatrix} \prod_{j=1}^{t} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P(Z^j|S_i^jC^j\pi_i) \\ \times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^jS_i^jB^{j-1}\pi_i) \\ \times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^jS_i^jB^j\pi_i) \\ \times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^jB^jS_i^jM^{j-1}\pi_i) \end{bmatrix} \\
\times P(M^0S_i^0C^0B^0Z_i^0\lambda_i^0\beta_i^0\alpha_i^0|\pi_{as}) \end{bmatrix}
\end{array} \right|
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \sum \begin{bmatrix} S^{0:t-1} \\ Z^t \\ C^t \\ B^t \\ M^t \\ \lambda^t \\ \beta^t \\ \alpha^t \end{bmatrix} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^t|S_i^t C^t \pi_i) \\
\times P(B^t|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^t|B^t S_i^t b^{t-1}\pi_i) \\
\times P(C^t|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^t|C^t S_i^t B^t \pi_i) \\
\times P(M^t|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^t|M^t B^t S_i^t \pi_i) \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{bmatrix} \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{bmatrix}
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \sum_{S^{0:t-1}} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_{as}) \\
\times \sum_{B^t} \begin{bmatrix}
P(B^t|\pi_{as}) \times \sum_{\beta^t} \prod_{i=1}^{N_i} P(\beta^t|B^t S_i^t b^{t-1}\pi_i) \\
\times \sum_{C^t} \begin{bmatrix} P(C^t|\pi_{as}) \times \sum_{\alpha^t} \prod_{i=1}^{N_i} P(\alpha^t|C^t S_i^t B^t \pi_i) \\ \times \sum_{Z^t} \prod_{i=1}^{N_i} P(Z^t|S_i^t C^t \pi_i) \end{bmatrix} \\
\times \sum_{M^t} \left[P(M^t|\pi_{as}) \times \sum_{\lambda^t} \prod_{i=1}^{N_i} P(\lambda^t|M^t B^t S_i^t \pi_i) \right] \\
\times
\end{bmatrix} \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{bmatrix} \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{bmatrix}
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \sum_{S^{0:t-1}} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{bmatrix} \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{bmatrix}
\end{array} \right|
\end{aligned}
\tag{D.9}
$$

D.2.3 Attention Selection

$$P(C^t B^t|z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as})$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as})} \\
\times \sum \begin{bmatrix} S^{0:t} \\ Z^t \\ M^t \\ \lambda^t \end{bmatrix} \begin{bmatrix} \prod_{j=1}^{t} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P(Z^j|S_i^jC^j\pi_i) \\ \times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^jS_i^jB^{j-1}\pi_i) \\ \times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^jS_i^jB^j\pi_i) \\ \times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^jB^jS_i^jM^{j-1}\pi_i) \end{bmatrix} \\ \times P(M^0S_i^0C^0B^0Z_i^0\lambda_i^0\beta_i^0\alpha_i^0|\pi_{as}) \end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as})} \\
\times \sum \begin{bmatrix} S^{0:t} \\ Z^t \\ M^t \\ \lambda^t \end{bmatrix} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P(Z^t|S_i^tC^t\pi_i) \\ \times P(B^t|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^t|B^tS_i^tb^{t-1}\pi_i) \\ \times P(C^t|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^t|C^tS_i^tB^t\pi_i) \\ \times P(M^t|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^t|M^tB^tS_i^t\pi_i) \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P(Z^j|S_i^jC^j\pi_i) \\ \times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^jS_i^jB^{j-1}\pi_i) \\ \times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^jS_i^jB^j\pi_i) \\ \times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^jB^jS_i^jM^{j-1}\pi_i) \end{bmatrix} \\ \times P(M^0S_i^0C^0B^0Z_i^0\lambda_i^0\beta_i^0\alpha_i^0|\pi_{as}) \end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as})} \\
\times \sum_{S^{0:t}} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\ \times \sum_{Z^t} \prod_{i=1}^{N_i} P(Z^t|S_i^tC^t\pi_i) \\ \times P(B^t|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^t|B^tS_i^tb^{t-1}\pi_i) \\ \times P(C^t|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^t|C^tS_i^tB^t\pi_i) \\ \times \sum_{M^t} \left[P(M^t|\pi_{as}) \times \sum_{\lambda^t} \prod_{i=1}^{N_i} P(\lambda^t|M^tB^tS_i^t\pi_i) \right] \\ \times \prod_{j=1}^{t-1} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P(Z^j|S_i^jC^j\pi_i) \\ \times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^jS_i^jB^{j-1}\pi_i) \\ \times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^jS_i^jB^j\pi_i) \\ \times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^jB^jS_i^jM^{j-1}\pi_i) \end{bmatrix} \\ \times P(M^0S_i^0C^0B^0Z_i^0\lambda_i^0\beta_i^0\alpha_i^0|\pi_{as}) \end{bmatrix}
\end{array} \right|
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as})} \\
\times \sum_{S^{0:t}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\times P(B^t|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^t|B^t S_i^t b^{t-1}\pi_i) \\
\times P(C^t|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^t|C^t S_i^t B^t \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as})} \\
\times \sum_{S^{t}} \left[\begin{array}{l}
P(B^t|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^t|B^t S_i^t b^{t-1}\pi_i) \\
\times P(C^t|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^t|C^t S_i^t B^t \pi_i) \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{array} \right]
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \times P(B^t|\pi_{as}) \times P(C^t|\pi_{as}) \\
\times \sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(\beta^t|B^t S_i^t b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(\alpha^t|C^t S_i^t B^t \pi_i) \\
\times P(S^t|z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}\pi_{as})
\end{array} \right]
\end{array} \right|
\end{aligned}
\tag{D.10}
$$

D.2.4 Behavior Selection

$$
\begin{aligned}
&P(B^t|z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as}) \\
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \\
\times \sum \left[\begin{array}{l} S^{0:t} \\ M^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \\
\times \sum \begin{bmatrix} S^{0:t} \\ M^t \\ \lambda^t \end{bmatrix} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(z^t|S_i^t c^t \pi_i) \\
\times P(B^t|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^t|B^t S_i^t b^{t-1}\pi_i) \\
\times P(c^t|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^t|c^t S_i^t B^t \pi_i) \\
\times P(M^t|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^t|M^t B^t S_i^t \pi_i) \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{bmatrix} \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \\
\times \sum_{S^{0:t}} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\times \prod_{i=1}^{N_i} P(z^t|S_i^t c^t \pi_i) \\
\times P(B^t|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^t|B^t S_i^t b^{t-1}\pi_i) \\
\times P(c^t|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^t|c^t S_i^t B^t \pi_i) \\
\times \sum_{M^t} \left[P(M^t|\pi_{as}) \times \sum_{\lambda^t} \prod_{i=1}^{N_i} P(\lambda^t|M^t B^t S_i^t \pi_i) \right] \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{bmatrix} \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{bmatrix}
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \\
\times \sum_{S^t} \begin{bmatrix}
P(B^t|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^t|B^t S_i^t b^{t-1}\pi_i) \\
\times P(c^t|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^t|c^t S_i^t B^t \pi_i) \\
\times \prod_{i=1}^{N_i} P(z^t|S_i^t c^t \pi_i) \\
\times \sum_{S^{0:t-1}} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\times \prod_{j=1}^{t-1} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{bmatrix} \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{bmatrix}
\end{bmatrix}
\end{array} \right|
$$

$$= \left| \begin{array}{l} \frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \times P(B^t|\pi_{as}) \times P(c^t|\pi_{as}) \\ \times \sum_{S^t} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(\beta^t|B^t S_i^t b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P(\alpha^t|c^t S_i^t B^t \pi_i) \\ \prod_{i=1}^{N_i} P(z^t|S_i^t c^t \pi_i) \\ \times P(S^t|z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}\pi_{as}) \end{array} \right] \end{array} \right| \tag{D.11}$$

D.2.5 Estimation

$$P(S^t|z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t}\pi_{as})$$

$$= \left| \begin{array}{l} \frac{1}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t}|\pi_{as})} \\ \times \sum \left[\begin{array}{l} S^{0:t-1} \\ B^t \\ M^t \\ \lambda^t \\ \beta^t \end{array} \right] \left[\begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\ \times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\ \times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\ \times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as}) \end{array} \right] \end{array} \right|$$

$$= \left| \begin{array}{l} \frac{1}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t}|\pi_{as})} \\ \times \sum \left[\begin{array}{l} S^{0:t-1} \\ B^t \\ M^t \\ \lambda^t \\ \beta^t \end{array} \right] \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P(z^t|S_i^t c^t \pi_i) \\ \times P(B^t|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^t|B^t S_i^t b^{t-1}\pi_i) \\ \times P(c^t|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^t|c^t S_i^t B^t \pi_i) \\ \times P(M^t|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^t|M^t B^t S_i^t \pi_i) \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\ \times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\ \times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\ \times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i) \end{array} \right] \\ \times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as}) \end{array} \right] \end{array} \right|$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t}|\pi_{as})} \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(z^t|S_i^t c^t \pi_i) \\
\times \sum_{B^t} \left[\begin{array}{l}
P(B^t|\pi_{as}) \times \sum_{\beta^t} \prod_{i=1}^{N_i} P(\beta^t|B^t S_i^t b^{t-1}\pi_i) \\
\times P(c^t|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^t|c^t S_i^t B^t \pi_i) \\
\times \sum_{M^t} \left[P(M^t|\pi_{as}) \times \sum_{\lambda^t} \prod_{i=1}^{N_i} P(\lambda^t|M^t B^t S_i^t \pi_i) \right]
\end{array} \right] \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t}|\pi_{as})} \times P(c^t|\pi_{as}) \\
\times \sum_{B^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(z^t|S_i^t c^t \pi_i) \\
\times \sum_{B^t} \left[\begin{array}{l}
P(B^t|\pi_{as}) \\
\prod_{i=1}^{N_i} P(\alpha^t|c^t S_i^t B^t \pi_i)
\end{array} \right] \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{array} \right]
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t}|\pi_{as})} \times P(c^t|\pi_{as}) \\
\times \sum_{B^t} \left[\begin{array}{l}
P(B^t|\pi_{as}) \\
\prod_{i=1}^{N_i} P(\alpha^t|c^t S_i^t B^t \pi_i)
\end{array} \right] \\
\times \prod_{i=1}^{N_i} P(z^t|S_i^t c^t \pi_i) \\
\times P(S^t|z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}\pi_{as})
\end{array} \right|
\end{aligned}
\tag{D.12}
$$

D.2.6 Motor Command

$$
\begin{aligned}
&P(M^t|z^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\alpha^{0:t}\pi_{as}) \\
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \\
\times \sum_{S^{0:t}} \left[\begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \\
\times \sum_{S^{0:t}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(z^t|S_i^t c^t \pi_i) \\
\times P(b^t|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^t|b^t S_i^t b^{t-1}\pi_i) \\
\times P(c^t|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^t|c^t S_i^t b^t \pi_i) \\
\times P(M^t|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^t|M^t b^t S_i^t \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(z^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \times P(b^t|\pi_{as}) \times P(c^t|\pi_{as}) \times P(M^t|\pi_{as}) \\
\times \sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(z^t|S_i^t c^t \pi_{as}) \\
\prod_{i=1}^{N_i} P(\beta^t|b^t S_i^t b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(\alpha^t|c^t S_i^t b^t \pi_i) \\
\prod_{i=1}^{N_i} P(\lambda^t|M^t b^t S_i^t \pi_i) \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^j C^j \pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^j S_i^j B^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^j S_i^j B^j \pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^j B^j S_i^j M^{j-1}\pi_i)
\end{array} \right] \\
\times P(M^0 S_i^0 C^0 B^0 Z_i^0 \lambda_i^0 \beta_i^0 \alpha_i^0|\pi_{as})
\end{array} \right]
\end{array} \right] \\
&= \left| \begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \times P(b^t|\pi_{as}) \times P(c^t|\pi_{as}) \times P(M^t|\pi_{as}) \\
\times \sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(z^t|S_i^t c^t \pi_i) \\
\prod_{i=1}^{N_i} P(\beta^t|b^t S_i^t b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(\alpha^t|c^t S_i^t b^t \pi_i) \\
\prod_{i=1}^{N_i} P(\lambda^t|M^t b^t S_i^t \pi_i) \\
\times P(S^t|z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}\pi_{as})
\end{array} \right]
\end{array} \right|
\end{aligned}
\tag{D.13}
$$

D.2.7 Prediction using Prediction

$$P(S^t|z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}\pi_{as})$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_{as}) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^jC^j\pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^jS_i^jB^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^jS_i^jB^j\pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^jB^jS_i^jM^{j-1}\pi_i)
\end{array} \right] \\
\times P(M^0S_i^0C^0B^0Z_i^0\lambda_i^0\beta_i^0\alpha_i^0|\pi_{as})
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(S_i^{t-1}|S_i^{t-2}m^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P(z^{t-1}|S_i^{t-1}c^{t-1}\pi_i) \\
\times P(b^{t-1}|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^{t-1}|b^{t-1}S_{i-1}^tb^{t-2}\pi_i) \\
\times P(c^{t-1}|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^{t-1}|c^{t-1}S_{i-1}^{t-1}b^{t-1}\pi_i) \\
\times P(m^{t-1}|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^{t-1}|m^{t-1}b^{t-1}S_i^{t-1}\pi_i) \\
\times \prod_{j=1}^{t-2} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^jC^j\pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^jS_i^jB^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^jS_i^jB^j\pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^jB^jS_i^jM^{j-1}\pi_i)
\end{array} \right] \\
\times P(M^0S_i^0C^0B^0Z_i^0\lambda_i^0\beta_i^0\alpha_i^0|\pi_{as})
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \times P(b^{t-1}|\pi_{as}) \times P(c^{t-1}|\pi_{as}) \times P(m^{t-1}|\pi_{as}) \\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(z^{t-1}|S_i^{t-1}c^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(\beta^{t-1}|b^{t-1}S_{i-1}^tb^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P(\alpha^{t-1}|c^{t-1}S_{i-1}^{t-1}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(\lambda^{t-1}|m^{t-1}b^{t-1}S_i^{t-1}\pi_i) \\
\times \sum_{S^{0:t-2}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^{t-1}|S_i^{t-2}m^{t-2}\pi_i) \\
\times \prod_{j=1}^{t-2} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j|S_i^{j-1}M^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P(Z^j|S_i^jC^j\pi_i) \\
\times P(B^j|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^j|B^jS_i^jB^{j-1}\pi_i) \\
\times P(C^j|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^j|C^jS_i^jB^j\pi_i) \\
\times P(M^j|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^j|M^jB^jS_i^jM^{j-1}\pi_i)
\end{array} \right] \\
\times P(M^0S_i^0C^0B^0Z_i^0\lambda_i^0\beta_i^0\alpha_i^0|\pi_{as})
\end{array} \right]
\end{array} \right]
\end{array} \right|
$$

$$= \left| \begin{array}{l} \frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \times P(b^{t-1}|\pi_{as}) \times P(c^{t-1}|\pi_{as}) \times P(m^{t-1}|\pi_{as}) \\ \times \sum_{S^{t-1}} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P(z^{t-1}|S_i^{t-1}c^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P(\beta^{t-1}|b^{t-1}S_{i-1}^t b^{t-2}\pi_i) \\ \prod_{i=1}^{N_i} P(\alpha^{t-1}|c^{t-1}S_{i-1}^{t-1}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P(\lambda^{t-1}|m^{t-1}b^{t-1}S_i^{t-1}\pi_i) \\ \times P(S^{t-1}|z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}\pi_{as}) \end{array} \right] \end{array} \right. \tag{D.14}$$

D.3 Relations Between Elementary and Global Filter

We aim to show that questions in global filter are proportional to the product of equivalent questions in elementary filter. So we want to show that:

$$\begin{aligned} &P(S^t|z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}\pi_{as}) \propto \\ &\quad \prod_{i=1}^{N_i} P(S_i^t|z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}\pi_i) \end{aligned} \tag{D.15}$$

$$\begin{aligned} &P(C^t|z^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t}\pi_{as}) \propto \\ &\quad \prod_{i=1}^{N_i} P(C^t|z_i^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t}\pi_i) \end{aligned} \tag{D.16}$$

$$\begin{aligned} &P(B^t|z^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as}) \propto \\ &\quad \prod_{i=1}^{N_i} P(B^t|z_i^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\alpha_i^{0:t}\pi_i) \end{aligned} \tag{D.17}$$

$$P(M^t|z^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\alpha^{0:t}\pi_{as}) \propto \tag{D.18}$$

$$\prod_{i=1}^{N_i} P(M^t|z_i^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t}\beta_i^{0:t}\alpha_i^{0:t}\pi_i) \tag{D.19}$$

Lets begin considering that prediction at time instant $t-1$ is proportional to elementary filters prediction:

$$\begin{aligned} &P(S^{t-1}|z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}\pi_{as}) \\ &\quad = \kappa_{t-2} \prod_{i=1}^{N_i} P(S_i^{t-1}|z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha^{0:t-2}\pi_i) \quad , \end{aligned} \tag{D.20}$$

where κ_{t-2} is a constant calculated from values of $z_i^{0:t-2}$, $c^{0:t-2}$, $b^{0:t-2}$, $m^{0:t-2}$, $\lambda_i^{0:t-2}$, $\beta_i^{0:t-2}$ and $\alpha^{0:t-2}$.

Applying equation above to calculate prediction at time instant t as stated in equation D.14, we have:

$$
\begin{aligned}
&P(S^t|z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}\pi_{as})\\
&= \left|\begin{array}{l}
\frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}\\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i)\\
\prod_{i=1}^{N_i} P(z^{t-1}|S_i^{t-1}c^{t-1}\pi_i)\\
\times P(b^{t-1}|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^{t-1}|b^{t-1}S_i^{t-1}b^{t-2}\pi_i)\\
\times P(c^{t-1}|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^{t-1}|c^{t-1}S_i^{t-1}b^{t-1}\pi_i)\\
\times P(m^{t-1}|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^{t-1}|m^{t-1}b^{t-1}S_i^{t-1}\pi_i)\\
\times P(S^{t-1}|z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}\pi_{as})
\end{array}\right]
\end{array}\right|\\
&= \left|\begin{array}{l}
\frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}\\
\times P(b^{t-1}|\pi_{as}) \times P(c^{t-1}|\pi_{as}) \times P(m^{t-1}|\pi_{as})\\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t|S_i^{t-1}m^{t-1}\pi_i)\\
\prod_{i=1}^{N_i} P(z^{t-1}|S_i^{t-1}c^{t-1}\pi_i)\\
\prod_{i=1}^{N_i} P(\beta^{t-1}|b^{t-1}S_i^{t-1}b^{t-2}\pi_i)\\
\prod_{i=1}^{N_i} P(\alpha^{t-1}|c^{t-1}S_i^{t-1}b^{t-1}\pi_i)\\
\prod_{i=1}^{N_i} P(\lambda^{t-1}|m^{t-1}b^{t-1}S_i^{t-1}\pi_i)\\
\times \kappa_{t-2} \prod_{i=1}^{N_i} P(S_i^{t-1}|z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha^{0:t-2}\pi_i)
\end{array}\right]
\end{array}\right|\\
&= \left|\begin{array}{l}
\frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}\\
\times P(b^{t-1}|\pi_{as}) \times P(c^{t-1}|\pi_{as}) \times P(m^{t-1}|\pi_{as})\\
\times \kappa_{t-2}\\
\times \sum_{S^{t-1}} \left[\prod_{i=1}^{N_i} \left[\begin{array}{l}
P(S_i^t|S_i^{t-1}m^{t-1}\pi_i)\\
\times P(z^{t-1}|S_i^{t-1}c^{t-1}\pi_i)\\
\times P(\beta^{t-1}|b^{t-1}S_i^{t-1}b^{t-2}\pi_i)\\
\times P(\alpha^{t-1}|c^{t-1}S_i^{t-1}b^{t-1}\pi_i)\\
\times P(\lambda^{t-1}|m^{t-1}b^{t-1}S_i^{t-1}\pi_i)\\
\times P(S_i^{t-1}|z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha^{0:t-2}\pi_i)
\end{array}\right]\right]
\end{array}\right|\\
&= \left|\begin{array}{l}
\frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}\\
\times \frac{P(b^{t-1}|\pi_{as})\times P(c^{t-1}|\pi_{as})\times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i)\times P(c^{t-1}|\pi_i)\times P(m^{t-1}|\pi_i)}\\
\times \kappa_{t-2}\\
\prod_{i=1}^{N_i} \left[\sum_{S_i^{t-1}} \left[\begin{array}{l}
P(S_i^t|S_i^{t-1}m^{t-1}\pi_i)\\
\times P(z^{t-1}|S_i^{t-1}c^{t-1}\pi_i)\\
\times P(b^{t-1}|\pi_i) \times P(\beta^{t-1}|b^{t-1}S_i^{t-1}b^{t-2}\pi_i)\\
\times P(c^{t-1}|\pi_i) \times P(\alpha^{t-1}|c^{t-1}S_i^{t-1}b^{t-1}\pi_i)\\
\times P(m^{t-1}|\pi_i) \times P(\lambda^{t-1}|m^{t-1}b^{t-1}S_i^{t-1}\pi_i)\\
\times P(S_i^{t-1}|z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha^{0:t-2}\pi_i)
\end{array}\right]\right]
\end{array}\right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \frac{P(b^{t-1}|\pi_{as})\times P(c^{t-1}|\pi_{as})\times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i)\times P(c^{t-1}|\pi_i)\times P(m^{t-1}|\pi_i)} \\
\times \kappa_{t-2} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)} \right] \\
\prod_{i=1}^{N_i} \left[\; P(S_i^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}\pi_i) \; \right]
\end{array} \right. \tag{D.21}
$$

From the equation above, we find the expression for κ_{t-1}:

$$
\kappa_{t-1} = \\
= \left| \begin{array}{l}
\frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \frac{P(b^{t-1}|\pi_{as})\times P(c^{t-1}|\pi_{as})\times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i)\times P(c^{t-1}|\pi_i)\times P(m^{t-1}|\pi_i)} \\
\times \kappa_{t-2} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)} \right]
\end{array} \right. \tag{D.22}
$$

It is necessary only verify now if:

$$
\begin{aligned}
& P(S^1|z^0b^0m^0\lambda^0\beta^0\alpha^0\pi_{as}) \\
& \quad = \left| \; \kappa_0 \prod_{i=1}^{N_i} P(S_i^1|z_i^0b^0m^0\lambda_i^0\beta_i^0\alpha_i^0\pi_i) \right.
\end{aligned} \tag{D.23}
$$

Verifiying expression for elementary filter, we have:

$$
\begin{aligned}
& P(S_i^1|z_i^0b^0m^0\lambda_i^0\beta_i^0\alpha_i^0\pi_i) \\
& = \left| \begin{array}{l}
\frac{1}{P(z_i^0b^0m^0\lambda_i^0\beta_i^0\alpha_i^0|\pi_i)} \times \\
\sum \begin{bmatrix} S_i^0 \\ M^1 \\ Z_i^1 \\ C^1 \\ B^1 \\ \lambda_i^1 \\ \beta_i^1 \\ \alpha_i^1 \end{bmatrix}
\begin{bmatrix}
P(M^0|\pi_i) \times P(S_i^0|\pi_i) \times P(C^0|\pi_i) \\
\times P(B^0|\pi_i) \times P(Z_i^0|S_i^0C^0\pi_i) \\
\times P(\lambda_i^0|M^0S^0\pi_i) \times P(\beta_i^0|B^0S^0\pi_i) \times P(\alpha_i^0|C^0B^0S^0\pi_i) \\
\times P(S_i^1|S_i^0m^0\pi_i) \\
\times P(Z_i^1|S_i^1C^1\pi_i) \\
\times P(B^1|\pi_i) \times P(\beta_i^1|B^1S_i^1B^0\pi_i) \\
\times P(C^1|\pi_i) \times P(\alpha_i^1|C^1S_i^1B^1\pi_i) \\
\times P(M^1|\pi_i) \times P(\lambda_i^1|M^tB^1S_i^1\pi_i)
\end{bmatrix}
\end{array} \right.
\end{aligned}
$$

$$= \left| \begin{array}{l} \frac{1}{P(z_i^0 b^0 m^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i)} \times \\ \sum_{S_i^0} \left[\begin{array}{l} P(M^0|\pi_i) \times P(S_i^0|\pi_i) \times P(C^0|\pi_i) \\ \times P(B^0|\pi_i) \times P(Z_i^0|S_i^0 C^0 \pi_i) \\ \times P(\lambda_i^0|M^0 S^0 \pi_i) \times P(\beta_i^0|B^0 S^0 \pi_i) \times P(\alpha_i^0|C^0 B^0 S^0 \pi_i) \\ \times P(S_i^1|S_i^0 m^0 \pi_i) \end{array} \right] \end{array} \right. \tag{D.24}$$

And for global filter we have:

$$P(S^1|z^0 b^0 m^0 \lambda^0 \beta^0 \alpha^0 \pi_{as})$$

$$= \left| \begin{array}{l} \frac{1}{P(z^0 b^0 m^0 \lambda^0 \beta^0 \alpha^0 | \pi_{as}))} \\ \times \sum \left[\begin{array}{l} S^0 \\ M^1 \\ Z^1 \\ C^1 \\ B^1 \\ \lambda^1 \\ \beta^1 \\ \alpha^1 \end{array} \right] \left[\begin{array}{l} P(M^0|\pi_{as}) \times P(S^0|\pi_{as}) \times P(C^0|\pi_{as}) \\ \times P(B^0|\pi_{as}) \times P(Z^0|S^0 C^0 \pi_{as}) \\ \times P(\lambda^0|M^0 S^0 B^0 \pi_{as}) \times P(\beta^0|B^0 S^0 \pi_{as}) \times P(\alpha^0|C^0 B^0 S^0 \pi_{as}) \\ \prod_{i=1}^{N_i} P(S_i^1|S_i^0 m^0 \pi_{as}) \\ \prod_{i=1}^{N_i} P(Z^1|S_i^1 C^1 \pi_{as}) \\ \times P(B^1|\pi_{as}) \prod_{i=1}^{N_i} P(\beta^1|B^1 S_i^1 B^0 \pi_i) \\ \times P(C^1|\pi_{as}) \prod_{i=1}^{N_i} P(\alpha^1|C^1 S_i^1 B^1 \pi_i) \\ \times P(M^1|\pi_{as}) \prod_{i=1}^{N_i} P(\lambda^1|M^1 B^1 S_i^1 \pi_i) \end{array} \right] \end{array} \right.$$

$$= \left| \begin{array}{l} \frac{1}{P(z^0 b^0 m^0 \lambda^0 \beta^0 \alpha^0 | \pi_{as}))} \\ \times P(M^0|\pi_{as}) \times P(C^0|\pi_{as}) \times P(B^0|\pi_{as}) \\ \times \sum_{S^0} \left[\begin{array}{l} P(S^0|\pi_{as}) \\ \times P(Z^0|S^0 C^0 \pi_{as}) \\ \times P(\lambda^0|M^0 S^0 B^0 \pi_{as}) \times P(\beta^0|B^0 S^0 \pi_{as}) \times P(\alpha^0|C^0 B^0 S^0 \pi_{as}) \\ \prod_{i=1}^{N_i} P(S_i^1|S_i^0 m^0 \pi_i) \\ \times \end{array} \right] \end{array} \right.$$

$$= \left| \begin{array}{l} \frac{1}{P(z^0 b^0 m^0 \lambda^0 \beta^0 \alpha^0 | \pi_{as}))} \\ \times P(M^0|\pi_{as}) \times P(C^0|\pi_{as}) \times P(B^0|\pi_{as}) \\ \times \sum_{S^0} \prod_{i=1}^{N_i} \left[\begin{array}{l} P(S_i^0|\pi_i) \\ \times P(Z_i^0|S_i^0 C^0 \pi_i) \\ \times P(\lambda_i^0|M^0 S_i^0 B^0 \pi_i) \\ \times P(\beta_i^0|B^0 S_i^0 \pi_i) \\ \times P(\alpha_i^0|C^0 B^0 S_i^0 \pi_i) \\ \times P(S_i^1|S_i^0 m^0 \pi_i) \end{array} \right] \end{array} \right.$$

$$
\begin{aligned}
&= \left| \begin{array}{l} \frac{1}{P(z^0 b^0 m^0 \lambda^0 \beta^0 \alpha^0 | \pi_{as}))} \\ \times \frac{P(M^0|\pi_{as}) \times P(C^0|\pi_{as}) \times P(B^0|\pi_{as})}{\prod_{i=1}^{N_i} [P(M^0|\pi_i) \times P(C^0|\pi_i) \times P(B^0|\pi_i)]} \\ \prod_{i=1}^{N_i} \sum_{S_i^0} \left[\begin{array}{l} P(S_i^0|\pi_i) \\ \times P(Z_i^0|S_i^0 C^0 \pi_i) \\ \times P(M^0|\pi_i) \times P(\lambda_i^0|M^0 S_i^0 B^0 \pi_i) \\ \times P(B^0|\pi_i) \times P(\beta_i^0|B^0 S_i^0 \pi_i) \\ \times P(C^0|\pi_i) \times P(\alpha_i^0|C^0 B^0 S_i^0 \pi_i) \\ \times P(S_i^1|S_i^0 m^0 \pi_i) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{\prod_{i=1}^{N_i} \left[P(z_i^0 b^0 m^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i) \right]}{P(z^0 b^0 m^0 \lambda^0 \beta^0 \alpha^0 | \pi_{as}))} \\ \times \frac{P(M^0|\pi_{as}) \times P(C^0|\pi_{as}) \times P(B^0|\pi_{as})}{\prod_{i=1}^{N_i} [P(M^0|\pi_i) \times P(C^0|\pi_i) \times P(B^0|\pi_i)]} \\ \times \prod_{i=1}^{N_i} P(S_i^1|z_i^0 b^0 m^0 \lambda_i^0 \beta_i^0 \alpha_i^0 \pi_i) \end{array} \right. \qquad \text{(D.25)}
\end{aligned}
$$

And so, we can find expression for κ_0:

$$
\begin{aligned}
&\kappa_0 \\
&= \left| \begin{array}{l} \frac{\prod_{i=1}^{N_i} \left[P(z_i^0 b^0 m^0 \lambda_i^0 \beta_i^0 \alpha_i^0 | \pi_i) \right]}{P(z^0 b^0 m^0 \lambda^0 \beta^0 \alpha^0 | \pi_{as}))} \\ \times \frac{P(M^0|\pi_{as}) \times P(C^0|\pi_{as}) \times P(B^0|\pi_{as})}{\prod_{i=1}^{N_i} [P(M^0|\pi_i) \times P(C^0|\pi_i) \times P(B^0|\pi_i)]} \end{array} \right. \qquad \text{(D.26)}
\end{aligned}
$$

Now, the relation between elementary filter attention question and global attention expression. Actually, demonstration for attention, behavior and motor questions is based on above demonstration of prediction relation. For brevity, only attention question will be examined, and procedure is equivalent for other questions.

From equation D.10, we have:

$$
\begin{aligned}
&P(C^t B^t | z^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t} \alpha^{0:t} \pi_{as}) \\
&= \left| \begin{array}{l} \frac{P(z^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t-1} \alpha^{0:t-1} | \pi_{as})}{P(z^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t} \alpha^{0:t} | \pi_{as})} \times P(B^t|\pi_{as}) \times P(C^t|\pi_{as}) \\ \times \sum_{S^t} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(\beta^t | B^t S_i^t b^{t-1} \pi_i) \\ \prod_{i=1}^{N_i} P(\alpha^t | C^t S_i^t B^t \pi_i) \\ \times P(S^t | z^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t-1} \alpha^{0:t-1} \pi_{as}) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(z^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t-1} \alpha^{0:t-1} | \pi_{as})}{P(z^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t} \alpha^{0:t} | \pi_{as})} \times P(B^t|\pi_{as}) \times P(C^t|\pi_{as}) \\ \times \sum_{S^t} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(\beta^t | B^t S_i^t b^{t-1} \pi_i) \\ \prod_{i=1}^{N_i} P(\alpha^t | C^t S_i^t B^t \pi_i) \\ \frac{P(z^{0:t-2} c^{0:t-2} b^{0:t-2} m^{0:t-2} \lambda^{0:t-2} \beta^{0:t-2} \alpha^{0:t-2} | \pi_{as})}{P(z^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda^{0:t-1} \beta^{0:t-1} \alpha^{0:t-1} | \pi_{as})} \\ \times \frac{P(b^{t-1}|\pi_{as}) \times P(c^{t-1}|\pi_{as}) \times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i) \times P(c^{t-1}|\pi_i) \times P(m^{t-1}|\pi_i)} \\ \times \kappa_{t-2} \\ \prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2} c^{0:t-2} b^{0:t-2} m^{0:t-2} \lambda_i^{0:t-2} \beta_i^{0:t-2} \alpha_i^{0:t-2} | \pi_i)}{P(z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} | \pi_i)} \right] \\ \times \prod_{i=1}^{N_i} \left[\; P(S_i^t | z_i^{0:t-1} c^{0:t-1} b^{0:t-1} m^{0:t-1} \lambda_i^{0:t-1} \beta_i^{0:t-1} \alpha_i^{0:t-1} \pi_i) \; \right] \end{array} \right] \end{array} \right.
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \times P(B^t|\pi_{as}) \times P(C^t|\pi_{as}) \\
\times \frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \frac{P(b^{t-1}|\pi_{as}) \times P(c^{t-1}|\pi_{as}) \times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i) \times P(c^{t-1}|\pi_i) \times P(m^{t-1}|\pi_i)} \\
\times \kappa_{t-2} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)} \right] \\
\times \sum_{S^t} \prod_{i=1}^{N_i} \left[\begin{array}{l} P(\beta^t|B^tS_i^tb^{t-1}\pi_i) \\ \times P(\alpha^t|C^tS_i^tB^t\pi_i) \\ \times P(S_i^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}\pi_i) \end{array} \right]
\end{array} \right.
$$

$$
= \left| \begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \times P(B^t|\pi_{as}) \times P(C^t|\pi_{as}) \\
\times \frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \frac{P(b^{t-1}|\pi_{as}) \times P(c^{t-1}|\pi_{as}) \times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i) \times P(c^{t-1}|\pi_i) \times P(m^{t-1}|\pi_i)} \\
\times \kappa_{t-2} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)} \right] \\
\prod_{i=1}^{N_i} \left[\begin{array}{l} P(\beta^t|B^tS_i^tb^{t-1}\pi_i) \\ \times P(\alpha^t|C^tS_i^tB^t\pi_i) \\ \times P(S_i^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}\pi_i) \end{array} \right]
\end{array} \right.
$$

Using equation D.3, we have:

$$
\begin{aligned}
& P(C^tB^t|z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as}) \\
& = \left| \begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \times P(B^t|\pi_{as}) \times P(C^t|\pi_{as}) \\
\times \frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \frac{P(b^{t-1}|\pi_{as}) \times P(c^{t-1}|\pi_{as}) \times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i) \times P(c^{t-1}|\pi_i) \times P(m^{t-1}|\pi_i)} \\
\times \kappa_{t-2} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)} \right] \\
\prod_{i=1}^{N_i} \frac{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\alpha_i^{0:t}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i) \times P(B^t|\pi_i) \times P(C^t|\pi_i)} \\
\prod_{i=1}^{N_i} P(C^tB^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\alpha_i^{0:t}\pi_i)
\end{array} \right.
\end{aligned}
\tag{D.27}
$$

Behavior selection is simpler.

$$
\begin{aligned}
& P(B^t|z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}\pi_{as}) \\
& = \left| \begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \times P(B^t|\pi_{as}) \times P(c^t|\pi_{as}) \\
\times \sum_{S^t} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(\beta_i^t|B^tS_i^tb^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P(\alpha_i^t|c^tS_i^tB^t\pi_i) \\ \prod_{i=1}^{N_i} P(z_i^t|S_i^tc^t\pi_i) \\ \times P(S^t|z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}\pi_{as}) \end{array} \right]
\end{array} \right.
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \times P(B^t|\pi_{as}) \times P(c^t|\pi_{as}) \\
\times \sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(\beta_i^t|B^t S_i^t b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(\alpha_i^t|c^t S_i^t B^t \pi_i) \\
\prod_{i=1}^{N_i} P(z_i^t|S_i^t c^t \pi_i) \\
\times \frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \frac{P(b^{t-1}|\pi_{as})\times P(c^{t-1}|\pi_{as})\times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i)\times P(c^{t-1}|\pi_i)\times P(m^{t-1}|\pi_i)} \\
\times \kappa_{t-2} \\
\times \prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)} \right] \\
\prod_{i=1}^{N_i} \left[\; P(S_i^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}\pi_i) \; \right]
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \times P(B^t|\pi_{as}) \times P(c^t|\pi_{as}) \\
\times \frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \frac{P(b^{t-1}|\pi_{as})\times P(c^{t-1}|\pi_{as})\times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i)\times P(c^{t-1}|\pi_i)\times P(m^{t-1}|\pi_i)} \\
\times \kappa_{t-2} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)} \right] \\
\times \sum_{S^t} \prod_{i=1}^{N_i} \left[\begin{array}{l}
P(\beta_i^t|B^t S_i^t b^{t-1}\pi_i) \\
\times P(\alpha_i^t|c^t S_i^t B^t \pi_i) \\
\times P(z_i^t|S_i^t c^t \pi_i) \\
\times P(S_i^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \times P(B^t|\pi_{as}) \times P(c^t|\pi_{as}) \\
\times \frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \frac{P(b^{t-1}|\pi_{as})\times P(c^{t-1}|\pi_{as})\times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i)\times P(c^{t-1}|\pi_i)\times P(m^{t-1}|\pi_i)} \\
\times \kappa_{t-2} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)} \right] \\
\prod_{i=1}^{N_i} \sum_{S_i^t} \left[\begin{array}{l}
P(\beta_i^t|B^t S_i^t b^{t-1}\pi_i) \\
\times P(\alpha_i^t|c^t S_i^t B^t \pi_i) \\
\times P(z_i^t|S_i^t c^t \pi_i) \\
\times P(S_i^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \times P(B^t|\pi_{as}) \times P(c^t|\pi_{as}) \\
\times \frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \frac{P(b^{t-1}|\pi_{as}) \times P(c^{t-1}|\pi_{as}) \times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i) \times P(c^{t-1}|\pi_i) \times P(m^{t-1}|\pi_i)} \\
\times \kappa_{t-2} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)} \right] \\
\times \frac{1}{\prod_{i=1}^{N_i} [P(B^t|\pi_i) \times P(c^t|\pi_i)]} \\
\times \frac{P(z_i^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\alpha_i^{0:t}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)} \\
\prod_{i=1}^{N_i} P(B^t|z_i^{0:t}c^{0:t}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t}\alpha_i^{0:t}\pi_i)
\end{array} \right. \tag{D.28}
$$

And for motor commands, a similar process is applied:

$$
\begin{aligned}
& P(M^t|z^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\alpha^{0:t}\pi_{as}) \\
&= \left| \begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \\
\times P(b^t|\pi_{as}) \times P(c^t|\pi_{as}) \times P(M^t|\pi_{as}) \\
\times \sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(z^t|S_i^t c^t \pi_i) \\
\times \prod_{i=1}^{N_i} P(\beta^t|b^t S_i^t b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(\alpha^t|c^t S_i^t b^t \pi_i) \\
\prod_{i=1}^{N_i} P(\lambda^t|M^t b^t S_i^t \pi_i) \\
\times P(S^t|z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}\pi_{as})
\end{array} \right]
\end{array} \right. \\
&= \left| \begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \\
\times P(b^t|\pi_{as}) \times P(c^t|\pi_{as}) \times P(M^t|\pi_{as}) \\
\times \sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(z_i^t|S_i^t c^t \pi_i) \\
\prod_{i=1}^{N_i} P(\beta_i^t|b^t S_i^t b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P(\alpha_i^t|c^t S_i^t b^t \pi_i) \\
\prod_{i=1}^{N_i} P(\lambda_i^t|M^t b^t S_i^t \pi_i) \\
\times \frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \frac{P(b^{t-1}|\pi_{as}) \times P(c^{t-1}|\pi_{as}) \times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i) \times P(c^{t-1}|\pi_i) \times P(m^{t-1}|\pi_i)} \\
\times \kappa_{t-2} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)} \right] \\
\prod_{i=1}^{N_i} \left[P(S_i^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}\pi_i) \right]
\end{array} \right]
\end{array} \right.
\end{aligned}
$$

$$
\begin{aligned}
&\left|\begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \\
\times P(b^t|\pi_{as}) \times P(c^t|\pi_{as}) \times P(M^t|\pi_{as}) \\
\times \frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \frac{P(b^{t-1}|\pi_{as})\times P(c^{t-1}|\pi_{as})\times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i)\times P(c^{t-1}|\pi_i)\times P(m^{t-1}|\pi_i)} \\
\times \kappa_{t-2}
\end{array}\right. \\
&= \left|\begin{array}{l}
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)}\right] \\
\times \sum_{S^t} \prod_{i=1}^{N_i} \left[\begin{array}{l}
P(z_i^t|S_i^t c^t \pi_i) \\
\times P(\beta_i^t|b^t S_i^t b^{t-1}\pi_i) \\
\times P(\alpha_i^t|c^t S_i^t b^t \pi_i) \\
\times P(\lambda_i^t|M^t b^t S_i^t \pi_i) \\
\times P(S_i^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}\pi_i)
\end{array}\right]
\end{array}\right. \\
&= \left|\begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \\
\times P(b^t|\pi_{as}) \times P(c^t|\pi_{as}) \times P(M^t|\pi_{as}) \\
\times \frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \frac{P(b^{t-1}|\pi_{as})\times P(c^{t-1}|\pi_{as})\times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i)\times P(c^{t-1}|\pi_i)\times P(m^{t-1}|\pi_i)} \\
\times \kappa_{t-2} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)}\right] \\
\prod_{i=1}^{N_i} \sum_{S_i^t} \left[\begin{array}{l}
P(z_i^t|S_i^t c^t \pi_i) \\
\times P(\beta_i^t|b^t S_i^t b^{t-1}\pi_i) \\
\times P(\alpha_i^t|c^t S_i^t b^t \pi_i) \\
\times P(\lambda_i^t|M^t b^t S_i^t \pi_i) \\
\times P(S_i^t|z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}\pi_i)
\end{array}\right]
\end{array}\right. \\
&= \left|\begin{array}{l}
\frac{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})}{P(z^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda^{0:t}\beta^{0:t}\alpha^{0:t}|\pi_{as})} \\
\times P(b^t|\pi_{as}) \times P(c^t|\pi_{as}) \times P(M^t|\pi_{as}) \\
\times \frac{P(z^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda^{0:t-2}\beta^{0:t-2}\alpha^{0:t-2}|\pi_{as})}{P(z^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda^{0:t-1}\beta^{0:t-1}\alpha^{0:t-1}|\pi_{as})} \\
\times \frac{P(b^{t-1}|\pi_{as})\times P(c^{t-1}|\pi_{as})\times P(m^{t-1}|\pi_{as})}{\prod_{i=1}^{N_i} P(b^{t-1}|\pi_i)\times P(c^{t-1}|\pi_i)\times P(m^{t-1}|\pi_i)} \\
\times \kappa_{t-2} \\
\prod_{i=1}^{N_i} \left[\frac{P(z_i^{0:t-2}c^{0:t-2}b^{0:t-2}m^{0:t-2}\lambda_i^{0:t-2}\beta_i^{0:t-2}\alpha_i^{0:t-2}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)}\right] \\
\times \frac{1}{\prod_{i=1}^{N_i} P(b^t|\pi_i)\times P(c^t|\pi_i)} \\
\times \frac{P(z_i^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t}\beta_i^{0:t}\alpha_i^{0:t}|\pi_i)}{P(z_i^{0:t-1}c^{0:t-1}b^{0:t-1}m^{0:t-1}\lambda_i^{0:t-1}\beta_i^{0:t-1}\alpha_i^{0:t-1}|\pi_i)} \\
\prod_{i=1}^{N_i} P(M^t|z_i^{0:t}c^{0:t}b^{0:t}m^{0:t-1}\lambda_i^{0:t}\beta_i^{0:t}\alpha_i^{0:t}\pi_i)
\end{array}\right.
\end{aligned}
\tag{D.29}
$$

Appendix E

Hierarchy of Behaviours

E.1 Elementary Filter

E.1.1 Joint Distribution

$$
\begin{aligned}
&P(S_i^{0:t} \wedge {}_{[1:L]}C^{0:t} \wedge {}_{[1:L]}Z_i^{0:t} \wedge {}_{[1:L]}B^{0:t} \wedge M^{0:t} \wedge \lambda_i^{0:t} \wedge {}_{[1:L]}\beta_i^{0:t} \wedge {}_{[1:L]}\alpha_i^{0:t} \,|\pi_i) \\
&= \left| \begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i)
\end{array} \right|
\end{aligned}
\tag{E.1}
$$

E.1.2 Prediction

In the following equations, prediction question for elementary filter is detailed and we define:

$$
\begin{aligned}
&P(S_i^t|\nu_p^t \wedge \pi_i) \\
&\quad = P(S_i^t|{}_{[1:L]}c^{0:t-1} \wedge {}_{[1:L]}\alpha_i^{0:t-1} \wedge {}_{[1:L]}z_i^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1:L]}\beta_i^{0:t-1} \wedge m^{0:t-1} \wedge \lambda_i^{0:t-1} \wedge \pi_i)
\end{aligned}
$$

Whenever right part of the question is too big, an auxiliar variable is employed for easiness of legibility. For the sake of brevity, $\nu_p^t \wedge \pi_i$ stands for right part of prediction

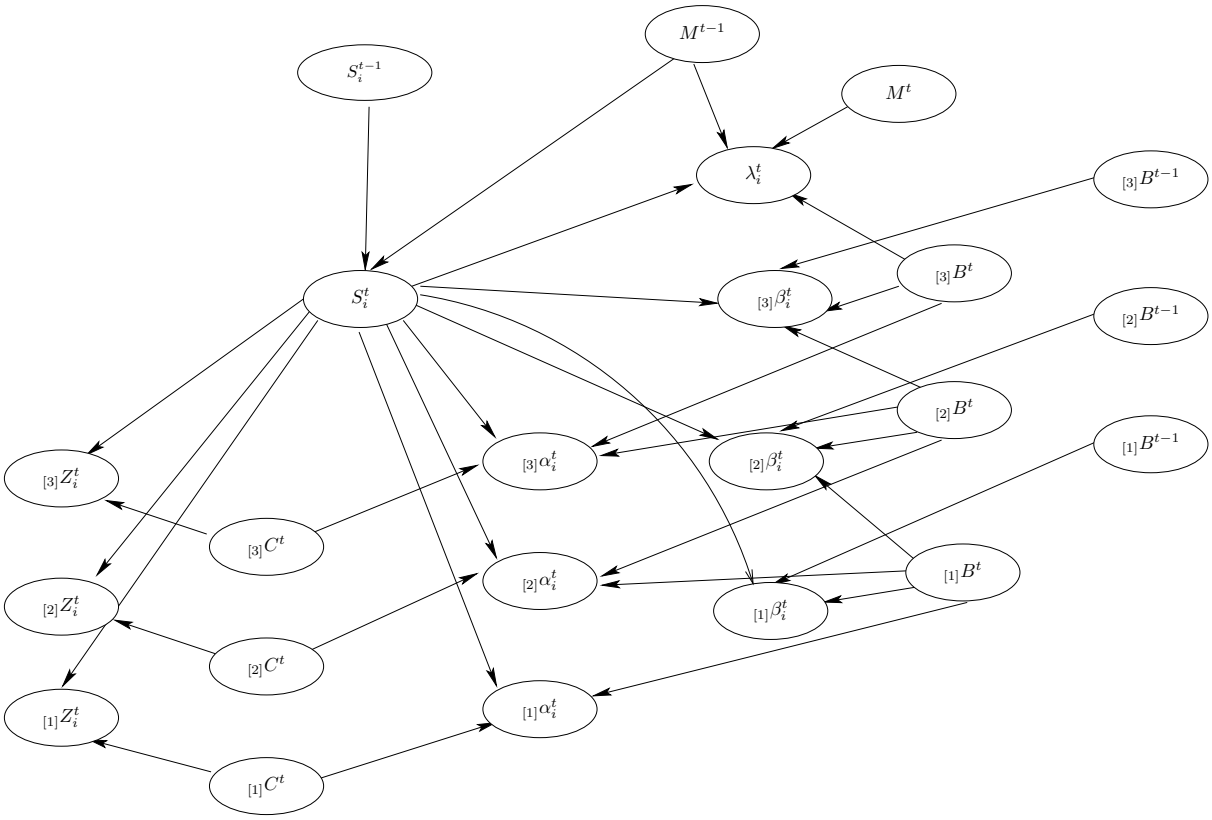


Figure E.1: Bayes Network for Elementary Filter defined in joint distribution E.1 for three levels (L=3).

question in elementary filter at time instant t, as $\nu_p^t \wedge \pi$ denotes right part of prediction question in global filter at time instant t.

$$P(S_i^t | \nu_p^t \wedge \pi_i)$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^t|\pi_i)} \\
\times \sum \left[\begin{array}{l} S_i^{0:t-1} \\ {}_{[1:L]}C^t \\ {}_{[1:L]}Z_i^t \\ {}_{[1:L]}B^t \\ M^t \\ {}_{[1:L]}\alpha_i^t \\ {}_{[1:L]}\beta_i^t \\ \lambda_i^t \end{array} \right]
\left[\begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^t|\pi_i)} \\
\times \sum \left[\begin{array}{l} S_i^{0:t-1} \\ {}_{[1:L]}C^t \\ {}_{[1:L]}Z_i^t \\ {}_{[1:L]}B^t \\ M^t \\ {}_{[1:L]}\alpha_i^t \\ {}_{[1:L]}\beta_i^t \\ \lambda_i^t \end{array} \right]
\left[\begin{array}{l}
P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\times P({}_{[1]}Z_i^t \,|S_i^t \wedge {}_{[1]}C^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^t|\pi_i)} \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times \sum \left[\begin{array}{l} {}_{[1:L]}C^t \\ {}_{[1:L]}Z_i^t \\ {}_{[1:L]}B^t \\ {}_{[1:L]}\alpha_i^t \\ {}_{[1:L]}\beta_i^t \\ M^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l}
P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\times P({}_{[1]}Z_i^t \,|S_i^t \wedge {}_{[1]}C^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right] \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i)
\end{array} \right]
\end{array} \right|
$$

It is easy to verify that the sum over variables at time step t sums to one.

$$
\sum_{\left[\begin{array}{l} {}_{[1:L]}C^t \\ {}_{[1:L]}Z_i^t \\ {}_{[1:L]}B^t \\ {}_{[1:L]}\alpha_i^t \\ {}_{[1:L]}\beta_i^t \\ M^t \\ \lambda_i^t \end{array} \right]} \left[\begin{array}{l}
P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\times P({}_{[1]}Z_i^t \,|S_i^t \wedge {}_{[1]}C^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right]
$$

$$
= \sum_{{}_{[1]}B^t} \left[\begin{array}{l}
P({}_{[1]}B^t|\pi_i) \times \sum_{{}_{[1]}\beta_i^t} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times \sum_{{}_{[1]}C^t} \left[\begin{array}{l} P({}_{[1]}C^t|\pi_i) \times \sum_{{}_{[1]}\alpha_i^t} P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \times \sum_{{}_{[1]}Z_i^t} \left[P({}_{[1]}Z_i^t|S_i^t \wedge {}_{[1]}C^t \wedge \pi_i) \right] \end{array} \right] \\
\times \sum_{{}_{[2:L]}B^t} \left[\begin{array}{l}
\prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi_i) \\ \times \sum_{{}_{[k]}\beta_i^t} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times \sum_{{}_{[k]}C^t} \left[\begin{array}{l} P({}_{[k]}C^t|\pi_i) \\ \times \sum_{{}_{[k]}\alpha_i^t} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \times \sum_{{}_{[k]}Z_i^t} P({}_{[k]}Z_i^t|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array} \right] \end{array} \right] \\
\times \sum_{M^t} \left[\begin{array}{l} P(M^t|\pi_i) \\ \times \sum_{\lambda_i^t} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array} \right]
\end{array} \right]
\end{array} \right]
$$

$$
P(S_i^t|\nu_p^t \wedge \pi_i) \tag{E.2}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^t|\pi_i)} \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t \left| S_i^{t-1} \wedge m^{t-1} \wedge \pi_i \right.) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \left| S_i^{j-1} \wedge M^{j-1} \wedge \pi_i \right.) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \left| S_i^j \wedge {}_{[1]}C^j \wedge \pi_i \right.) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \left| S_i^j \wedge {}_{[k]}C^j \wedge \pi_i \right.)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right]
\end{array} \right.
$$

Expressing prediction as function of previous time steps, we have:

$$
P(S_i^t|\nu_p^t \wedge \pi_i)
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^t|\pi_i)} \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P(S_i^{t-1} \,|\, S_i^{t-2} \wedge m^{t-2} \wedge \pi_i) \\
\times P({}_{[1]}b^{t-1}|\pi_i) \times P({}_{[1]}c^{t-1}|\pi_i) \\
\times P({}_{[1]}\beta_i^{t-1}|{}_{[1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-2}\pi_i) \\
\times P({}_{[1]}\alpha_i^{t-1}|{}_{[1]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[1]}c^{t-1} \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^{t-1}|\pi_i) \times P({}_{[k]}c^{t-1}|\pi_i) \\
\times P({}_{[k]}\beta_i^{t-1}|{}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-2}\pi_i) \\
\times P({}_{[k]}\alpha_i^{t-1}|{}_{[k]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1}\pi_i) \\
\times P({}_{[k]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[k]}c^{t-1} \wedge \pi_i)
\end{array} \right] \\
\times P(m^{t-1}|\pi_i) \times P(\lambda_i^{t-1}|m^{t-1} \wedge m^{t-2} \wedge {}_{[L]}b^{t-1} \wedge S_i^{t-1}\pi_i) \\
\prod_{j=1}^{t-2} \left[\begin{array}{l}
P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{1}{P(\nu_p^t|\pi_i)} \\
\times \sum_{S_i^{t-1}} \left[\begin{array}{l}
P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^{t-1}|\pi_i) \times P({}_{[1]}c^{t-1}|\pi_i) \\
\times P({}_{[1]}\beta_i^{t-1}|{}_{[1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-2}\pi_i) \\
\times P({}_{[1]}\alpha_i^{t-1}|{}_{[1]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[1]}c^{t-1} \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^{t-1}|\pi_i) \times P({}_{[k]}c^{t-1}|\pi_i) \\
\times P({}_{[k]}\beta_i^{t-1}|{}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-2}\pi_i) \\
\times P({}_{[k]}\alpha_i^{t-1}|{}_{[k]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1}\pi_i) \\
\times P({}_{[k]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[k]}c^{t-1} \wedge \pi_i)
\end{array} \right] \\
\times P(m^{t-1}|\pi_i) \times P(\lambda_i^{t-1}|m^{t-1} \wedge m^{t-2} \wedge {}_{[L]}b^{t-1} \wedge S_i^{t-1}\pi_i) \\
\times \sum_{S_i^{0:t-2}} \left[\begin{array}{l}
P(S_i^{t-1} \,|\, S_i^{t-2} \wedge m^{t-2} \wedge \pi_i) \\
\prod_{j=1}^{t-2} \left[\begin{array}{l}
P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]} b^0{}_{[1:L]} c^0{}_{[1:L]} z_i^0{}_{[0:L]} \beta_i^0{}_{[1:L]} \alpha_i^0 \lambda_i^0 \pi_i)
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(\nu_p^{t-1}|\pi_i)}{P(\nu_p^t|\pi_i)} \\
\times \sum_{S_i^{t-1}} \left[\begin{array}{l}
P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^{t-1}|\pi_i) \times P({}_{[1]}c^{t-1}|\pi_i) \times \\
P({}_{[1]}\beta_i^{t-1}|{}_{[1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-2}\pi_i) \\
\times P({}_{[1]}\alpha_i^{t-1}|{}_{[1]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[1]}c^{t-1} \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^{t-1}|\pi_i) \times P({}_{[k]}c^{t-1}|\pi_i) \\
\times P({}_{[k]}\beta_i^{t-1}|{}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-2}\pi_i) \\
\times P({}_{[k]}\alpha_i^{t-1}|{}_{[k]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1}\pi_i) \\
\times P({}_{[k]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[k]}c^{t-1} \wedge \pi_i)
\end{array} \right] \\
\times P(m^{t-1}|\pi_i) \times P(\lambda_i^{t-1}|m^{t-1} \wedge m^{t-2} \wedge {}_{[L]}b^{t-1} \wedge S_i^{t-1}\pi_i) \\
\times P(S_i^{t-1}|\nu_p^{t-1} \wedge \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
\tag{E.3}
$$

E.1.3 Attention Selection

Attention at level 1

Set of known variable values for attention question at level 1 is defined by variable ${}_{[1]}\nu_a^t$:

$$ {}_{[1]}\nu_a^t \wedge \pi_i = {}_{[1:L]}c^{0:t-1} \wedge {}_{[1:L]}\alpha_i^{0:t} \wedge {}_{[1:L]}z_i^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1:L]}\beta_i^{0:t} \wedge m^{0:t-1} \wedge \lambda_i^{0:t-1}. $$

$$ \begin{aligned} &P({}_{[1]}C^t \wedge {}_{[1]}B^t|{}_{[1]}\nu_a^t \wedge \pi_i) = \\ &\quad = P({}_{[1]}C^t \wedge {}_{[1]}B^t|{}_{[1:L]}c^{0:t-1} \wedge {}_{[1:L]}\alpha_i^{0:t} \wedge {}_{[1:L]}z_i^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1:L]}\beta_i^{0:t} \wedge m^{0:t-1} \wedge \lambda_i^{0:t-1} \wedge \pi_i) \end{aligned} $$

$$ \begin{aligned} &P({}_{[1]}C^t \wedge {}_{[1]}B^t|{}_{[1]}\nu_a^t \wedge \pi_i) \\ &= \left| \begin{array}{l} \frac{1}{P({}_{[1]}\nu_a^t|\pi_i)} \\ \times \sum \begin{bmatrix} S_i^{0:t} \\ {}_{[2:L]}C^t \\ {}_{[1:L]}Z_i^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha_i^t \\ {}_{[2:L]}\beta_i^t \\ \lambda_i^t \end{bmatrix} \left[\begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\ \times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\ \times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\ \times P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\ \times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \times P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{array} \right] \\ \times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i) \end{array} \right] \\ \times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i) \end{array} \right] \end{array} \right| \end{aligned} $$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_a^t|\pi_i)} \\
\times \sum \left[\begin{array}{l} S_i^{0:t} \\ {}_{[2:L]}C^t \\ {}_{[1:L]}Z_i^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha_i^t \\ {}_{[2:L]}\beta_i^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l}
P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\times P({}_{[1]}Z_i^t \,|S_i^t \wedge {}_{[1]}C^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_a^t|\pi_i)} \\
\times \sum_{S_i^{0:t}} \left[\begin{array}{l}
P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\times \sum_{{}_{[1]}Z_i^t} \left[P({}_{[1]}Z_i^t \,|S_i^t \wedge {}_{[1]}C^t \wedge \pi_i) \right] \\
\times \sum \left[\begin{array}{l} {}_{[2:L]}C^t \\ {}_{[2:L]}Z_i^t \\ {}_{[2:L]}B^t \\ {}_{[2:L]}\alpha_i^t \\ {}_{[2:L]}\beta_i^t \\ M^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l} \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\ \times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array} \right] \\ \times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array} \right] \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\ \times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right] \right|
$$

The sum over variables at time t is equal to one, as shown below.

$$
\sum_{\left[\begin{array}{l} {}_{[2:L]}C^t \\ {}_{[2:L]}Z_i^t \\ {}_{[2:L]}B^t \\ {}_{[2:L]}\alpha_i^t \\ {}_{[2:L]}\beta_i^t \\ M^t \\ \lambda_i^t \end{array} \right]} \left[\begin{array}{l} \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\ \times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array} \right] \\ \times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array} \right]
$$

$$
= \sum_{{}_{[2:L]}B^t} \left[\begin{array}{l} \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi_i) \\ \times \sum_{{}_{[k]}\beta_i^t} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times \sum_{{}_{[k]}C^t} \left[\begin{array}{l} P({}_{[k]}C^t|\pi_i) \\ \times \sum_{{}_{[k]}\alpha_i^t} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \times \sum_{{}_{[k]}Z_i^t} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array} \right] \end{array} \right] \\ \times \sum_{M^t} \left[\begin{array}{l} P(M^t|\pi_i) \\ \times \sum_{\lambda_i^t} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array} \right] \end{array} \right]
$$

And we can see that attention question is function of prediction.

$$
\begin{aligned}
& P({}_{[1]}C^t \wedge {}_{[1]}B^t|{}_{[1]}\nu_a^t \wedge \pi_i) \\
&= \left| \begin{array}{l} \frac{1}{P({}_{[1]}\nu_a^t|\pi_i)} \\ \times \sum_{S_i^t} \left[\begin{array}{l} P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i) \\ \times P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \times P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \times \sum_{S_i^{0:t-1}} \left[\begin{array}{l} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\ \times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\ \times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\ \times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\ \times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{array} \right] \\ \times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i) \end{array} \right] \\ \times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i) \end{array} \right] \end{array} \right] \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi_i)}{P({}_{[1]}\nu_a^t|\pi_i)} \\ \times \sum_{S_i^t} \left[\begin{array}{l} P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i) \\ \times P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \times P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \times P(S_i^t|\nu_p^t \wedge \pi_i) \end{array} \right] \end{array} \right.
\end{aligned}
\tag{E.4}
$$

Attention at level h, with $2 \leq h \leq L$

Here, right part of the question includes variables from upper levels that are already known: ${}_{[1:h-1]}c^{0:t}$ and ${}_{[1:h]}\alpha_i^{0:t}$, for example.

$$
{}_{[h]}\nu_a^t \wedge \pi_i = \left| \begin{array}{l} {}_{[1:h-1]}c^{0:t} \wedge {}_{[1:h]}\alpha_i^{0:t} \wedge {}_{[h:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha_i^{0:t-1} \wedge \\ {}_{[1:h-1]}b^{0:t} \wedge {}_{[1:h]}\beta_i^{0:t} \wedge {}_{[h:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta_i^{0:t-1} \wedge \\ {}_{[1:h-1]}z_i^{0:t} \wedge {}_{[h:L]}z_i^{0:t-1} \wedge m^{0:t-1} \wedge \lambda_i^{0:t-1} \end{array} \right.
$$

$$
\begin{aligned}
&P({}_{[h]}C^t \wedge {}_{[h]}B^t | {}_{[h]}\nu_a^t \wedge \pi_i) \\
&\quad = P({}_{[h]}C^t \wedge {}_{[h]}B^t | \begin{bmatrix} {}_{[1:h-1]}c^{0:t} \wedge {}_{[1:h]}\alpha_i^{0:t} \wedge {}_{[h:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha_i^{0:t-1} \wedge \\ {}_{[1:h-1]}b^{0:t} \wedge {}_{[1:h]}\beta_i^{0:t} \wedge {}_{[h:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta_i^{0:t-1} \wedge \\ {}_{[1:h-1]}z_i^{0:t} \wedge {}_{[h:L]}z_i^{0:t-1} \wedge m^{0:t-1} \wedge \lambda_i^{0:t-1} \wedge \pi_i \end{bmatrix})
\end{aligned}
$$

$$
\begin{aligned}
&P({}_{[h]}C^t \wedge {}_{[h]}B^t | {}_{[h]}\nu_a^t \wedge \pi_i) \\
&= \left| \begin{array}{l} \frac{1}{P({}_{[h]}\nu_a^t | \pi_i)} \\ \times \sum \begin{bmatrix} S_i^{0:t} \\ {}_{[h+1:L]}C^t \\ {}_{[h:L]}Z_i^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha_i^t \\ {}_{[h+1:L]}\beta_i^t \\ \lambda_i^t \end{bmatrix} \begin{bmatrix} \prod_{j=1}^{t} \begin{bmatrix} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j | \pi_i) \times P({}_{[1]}C^j | \pi_i) \\ \times P({}_{[1]}\beta_i^j | {}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1} \pi_i) \\ \times P({}_{[1]}\alpha_i^j | {}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j \pi_i) \\ \times P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \begin{bmatrix} P({}_{[k]}B^j | \pi_i) \times P({}_{[k]}C^j | \pi_i) \\ \times P({}_{[k]}\beta_i^j | {}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1} \pi_i) \\ \times P({}_{[k]}\alpha_i^j | {}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j \pi_i) \\ \times P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{bmatrix} \\ \times P(M^j | \pi_i) \times P(\lambda_i^j | M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j \pi_i) \end{bmatrix} \\ \times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i) \end{bmatrix} \end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_a^t|\pi_i)} \\
\times \sum \begin{bmatrix} S_i^{0:t} \\ {}_{[h+1:L]}C^t \\ {}_{[h:L]}Z_i^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha_i^t \\ {}_{[h+1:L]}\beta_i^t \\ \lambda_i^t \end{bmatrix}
\end{array} \right.
\left[\begin{array}{l}
P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h-1} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P({}_{[h]}B^t|\pi_i) \times P({}_{[h]}C^t|\pi_i) \\
\times P({}_{[h]}\beta_i^t|{}_{[h]}B^t \wedge {}_{[h-1]}b^t \wedge S_i^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\times P({}_{[h]}\alpha_i^t|{}_{[h]}C^t \wedge S_i^t \wedge {}_{[h]}B^t \wedge {}_{[h-1]}b^t\pi_i) \\
\times P({}_{[h]}Z_i^t \,|S_i^t \wedge {}_{[h]}C^t \wedge \pi_i) \\
\prod_{k=h+1}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i)
\end{array} \right]
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_a^t|\pi_i)} \\
\times \sum_{S_i^{0:t}} \left[\begin{array}{l}
P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h-1} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P({}_{[h]}B^t|\pi_i) \times P({}_{[h]}C^t|\pi_i) \\
\times P({}_{[h]}\beta_i^t|{}_{[h]}B^t \wedge {}_{[h-1]}b^t \wedge S_i^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\times P({}_{[h]}\alpha_i^t|{}_{[h]}C^t \wedge S_i^t \wedge {}_{[h]}B^t \wedge {}_{[h-1]}b^t\pi_i) \\
\times \sum_{{}_{[h]}Z_i^t} P({}_{[h]}Z_i^t \,|S_i^t \wedge {}_{[h]}C^t \wedge \pi_i) \\
\times \sum \left[\begin{array}{l}
{}_{[h+1:L]}C^t \\
{}_{[h+1:L]}Z_i^t \\
{}_{[h+1:L]}B^t \\
{}_{[h+1:L]}\alpha_i^t \\
{}_{[h+1:L]}\beta_i^t \\
M^t \\
\lambda_i^t
\end{array} \right] \prod_{k=h+1}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right] \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right]
\end{array} \right|
$$

It is possible to show that sum over variables concerning lower levels in the hierarchy is equal to one.

$$\sum_{\begin{bmatrix} {}_{[h+1:L]}C^t \\ {}_{[h+1:L]}Z_i^t \\ {}_{[h+1:L]}B^t \\ {}_{[h+1:L]}\alpha_i^t \\ {}_{[h+1:L]}\beta_i^t \\ M^t \\ \lambda_i^t \end{bmatrix}} \left\{ \prod_{k=h+1}^{L} \begin{bmatrix} P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\ \times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \times P({}_{[k]}Z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \\ \times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{bmatrix} \right\}$$

$$= \sum_{{}_{[h+1:L]}B^t} \left\{ \prod_{k=h+1}^{L} \begin{bmatrix} P({}_{[k]}B^t|\pi_i) \\ \times \sum_{{}_{[k]}\beta_i^t} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times \sum_{{}_{[k]}C^t} \begin{bmatrix} P({}_{[k]}C^t|\pi_i) \\ \times \sum_{{}_{[k]}\alpha_i^t} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \times \sum_{{}_{[k]}Z_i^t} P({}_{[k]}Z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \\ \times \end{bmatrix} \times \\ \sum_{M^t} \begin{bmatrix} P(M^t|\pi_i) \\ \times \sum_{\lambda_i^t} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{bmatrix} \end{bmatrix} \right\}$$

And attention is function of prediction question, as can be seen below.

$$P({}_{[h]}C^t \wedge {}_{[h]}B^t|{}_{[h]}\nu_a^t \wedge \pi_i)$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_a^t|\pi_i)} \\
\times \sum_{S_i^{0:t}} \left[\begin{array}{l}
P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h-1} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P({}_{[h]}B^t|\pi_i) \times P({}_{[h]}C^t|\pi_i) \\
\times P({}_{[h]}\beta_i^t|_{[h]}B^t \wedge {}_{[h-1]}b^t \wedge S_i^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\times P({}_{[h]}\alpha_i^t|_{[h]}C^t \wedge S_i^t \wedge {}_{[h]}B^t \wedge {}_{[h-1]}b^t\pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_a^t|\pi_i)} \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h-1} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P({}_{[h]}B^t|\pi_i) \times P({}_{[h]}C^t|\pi_i) \\
\times P({}_{[h]}\beta_i^t|{}_{[h]}B^t \wedge {}_{[h-1]}b^t \wedge S_i^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\times P({}_{[h]}\alpha_i^t|{}_{[h]}C^t \wedge S_i^t \wedge {}_{[h]}B^t \wedge {}_{[h-1]}b^t\pi_i) \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi_i)}{P({}_{[h]}\nu_a^t|\pi_i)} \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h-1} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P({}_{[h]}B^t|\pi_i) \times P({}_{[h]}C^t|\pi_i) \\
\times P({}_{[h]}\beta_i^t|{}_{[h]}B^t \wedge {}_{[h-1]}b^t \wedge S_i^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\times P({}_{[h]}\alpha_i^t|{}_{[h]}C^t \wedge S_i^t \wedge {}_{[h]}B^t \wedge {}_{[h-1]}b^t\pi_i) \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array} \right]
\end{array} \right|
$$

E.1.4 Behavior Selection

Action at level 1

Knowledge about attention variable at this level is added in the right part of the behaviour question, in relation to right part in the attention question.

$$
{}_{[1]}\nu_b^t \wedge \pi_i = \left| \begin{array}{l} {}_{[1]}c^{0:t} \wedge {}_{[1]}\alpha_i^{0:t} \wedge {}_{[2:L]}c^{0:t-1} \wedge {}_{[2:L]}\alpha_i^{0:t-1} \wedge \\ {}_{[1]}z_i^{0:t} \wedge {}_{[2:L]}z_i^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1]}\beta_i^{0:t} \wedge {}_{[2:L]}\beta_i^{0:t-1} \wedge \\ m^{0:t-1} \wedge \lambda_i^{0:t-1} \wedge \pi_i \end{array} \right.
$$

$$
\begin{aligned}
&P({}_{[1]}B^t|{}_{[1]}\nu_b^t \wedge \pi_i) \\
&\quad = P({}_{[1]}B^t| \left[\begin{array}{l} {}_{[1]}c^{0:t} \wedge {}_{[1]}\alpha_i^{0:t} \wedge {}_{[2:L]}c^{0:t-1} \wedge {}_{[2:L]}\alpha_i^{0:t-1} \wedge \\ {}_{[1]}z_i^{0:t} \wedge {}_{[2:L]}z_i^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1]}\beta_i^{0:t} \wedge {}_{[2:L]}\beta_i^{0:t-1} \wedge \\ m^{0:t-1} \wedge \lambda_i^{0:t-1} \wedge \pi_i \end{array} \right])
\end{aligned}
$$

$$
\begin{aligned}
&P({}_{[1]}B^t|{}_{[1]}\nu_b^t \wedge \pi_i) \\
&= \left| \begin{array}{l} \frac{1}{P({}_{[1]}\nu_b^t|\pi_i)} \\ \times \sum \left[\begin{array}{l} S_i^{0:t} \\ {}_{[2:L]}C^t \\ {}_{[2:L]}Z_i^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha_i^t \\ {}_{[2:L]}\beta_i^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\ \times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\ \times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\ \times P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\ \times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \times P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{array} \right] \\ \times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i) \end{array} \right] \\ \times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i) \end{array} \right] \end{array} \right.
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_b^t|\pi_i)} \\
\times \sum \left[\begin{array}{l} S_i^{0:t} \\ {}_{[2:L]}C^t \\ {}_{[2:L]}Z_i^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha_i^t \\ {}_{[2:L]}\beta_i^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l}
P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_b^t|\pi_i)} \\
\times \sum_{S_i^{0:t}} \left[\begin{array}{l}
P(S_i^t \,\big|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\times \sum \left[\begin{array}{l} {}_{[2:L]}C^t \\ {}_{[2:L]}Z_i^t \\ {}_{[2:L]}B^t \\ {}_{[2:L]}\alpha_i^t \\ {}_{[2:L]}\beta_i^t \\ M^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l} \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\ \times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array} \right] \\ \times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array} \right] \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,\big|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,\big|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\ \times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \times P({}_{[k]}Z_i^j \,\big|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i)
\end{array} \right] \right|
$$

Summing over variables at lower levels gives a result equal to one, as shown below.

$$
\sum_{\left[\begin{array}{l} {}_{[2:L]}C^t \\ {}_{[2:L]}Z_i^t \\ {}_{[2:L]}B^t \\ {}_{[2:L]}\alpha_i^t \\ {}_{[2:L]}\beta_i^t \\ M^t \\ \lambda_i^t \end{array} \right]} \left[\begin{array}{l} \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\ \times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array} \right] \\ \times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array} \right]
$$

$$= \sum_{[2:L]B^t} \left[\begin{array}{l} \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi_i) \times \sum_{{}_{[k]}\beta_i^t} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times \sum_{{}_{[k]}C^t} \left[\begin{array}{l} P({}_{[k]}C^t|\pi_i) \\ \times \sum_{{}_{[k]}\alpha_i^t} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \times \sum_{{}_{[k]}Z_i^t} P({}_{[k]}Z_i^t|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array} \right] \end{array} \right] \\ \times \sum_{M^t} \left[\begin{array}{l} P(M^t|\pi_i) \\ \times \sum_{\lambda_i^t} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array} \right] \end{array} \right]$$

Behaviour question can also be written in function of prediction expression.

$$P({}_{[1]}B^t|{}_{[1]}\nu_b^t \wedge \pi_i)$$

$$= \left| \begin{array}{l} \frac{1}{P({}_{[1]}\nu_b^t|\pi_i)} \\ \times \sum_{S_i^{0:t}} \left[\begin{array}{l} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\ \times P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\ \times P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \times P({}_{[1]}z_i^t|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\ \times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\ \times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\ \times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\ \times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{array} \right] \\ \times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i) \end{array} \right] \\ \times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i) \end{array} \right] \end{array} \right|$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_b^t|\pi_i)} \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t \,\big|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,\big|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,\big|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,\big|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi_i)}{P({}_{[1]}\nu_b^t|\pi_i)} \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array} \right]
\end{array} \right| \tag{E.5}
$$

Action at level $2 \leq h \leq L$

$$
{}_{[h]}\nu_b^t \wedge \pi_i = \left| \begin{array}{l}
{}_{[1:h]}c^{0:t} \wedge {}_{[1:h]}\alpha_i^{0:t} \wedge {}_{[h+1:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha_i^{0:t-1} \wedge \\
{}_{[1:h-1]}b^{0:t} \wedge {}_{[1:h]}\beta_i^{0:t} \wedge {}_{[h+1:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta_i^{0:t-1} \wedge \\
{}_{[1:h]}z_i^{0:t} \wedge {}_{[h+1:L]}z_i^{0:t-1} \wedge m^{0:t-1} \wedge \lambda_i^{0:t-1} \wedge \pi_i
\end{array} \right.
$$

$$
\begin{aligned}
& P({}_{[h]}B^t|_{[h]}\nu_b^t \wedge \pi_i) \\
& \quad = P({}_{[h]}B^t| \left[\begin{array}{l}
{}_{[1:h]}c^{0:t} \wedge {}_{[1:h]}\alpha_i^{0:t} \wedge {}_{[h+1:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha_i^{0:t-1} \wedge \\
{}_{[1:h-1]}b^{0:t} \wedge {}_{[1:h]}\beta_i^{0:t} \wedge {}_{[h+1:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta_i^{0:t-1} \wedge \\
{}_{[1:h]}z_i^{0:t} \wedge {}_{[h+1:L]}z_i^{0:t-1} \wedge m^{0:t-1} \wedge \lambda_i^{0:t-1} \wedge \pi_i
\end{array} \right])
\end{aligned}
$$

$$
P({}_{[h]}B^t|_{[h]}\nu_b^t \wedge \pi_i)
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_b^t|\pi_i)} \\
\times \sum \left[\begin{array}{l} S_i^{0:t} \\ {}_{[h+1:L]}C^t \\ {}_{[h+1:L]}Z_i^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha_i^t \\ {}_{[h+1:L]}\beta_i^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_b^t|\pi_i)} \\
\times \sum \left[\begin{array}{l} S_i^{0:t} \\ {}_{[h+1:L]}C^t \\ {}_{[h+1:L]}Z_i^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha_i^t \\ {}_{[h+1:L]}\beta_i^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l}
P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\prod_{k=h+1}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_b^t|\pi_i)} \\
\times \sum_{S_i^{0:t}} \left[\begin{array}{l}
P(S_i^t \,\big|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times \sum \left[\begin{array}{l}
{}_{[h+1:L]}C^t \\
{}_{[h+1:L]}Z_i^t \\
{}_{[h+1:L]}B^t \\
{}_{[h+1:L]}\alpha_i^t \\
{}_{[h+1:L]}\beta_i^t \\
M^t \\
\lambda_i^t
\end{array} \right] \prod_{k=h+1}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right] \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,\big|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,\big|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,\big|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_b^t|\pi_i)} \\
\times \sum_{S_i^{0:t}} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right] \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_b^t|\pi_i)} \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi_i)}{P({}_{[h]}\nu_b^t|\pi_i)} \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array} \right]
\end{array} \right|
$$

E.1.5 Estimation

Estimation at level 1

$$
{}_{[1]}\nu_e^t \wedge \pi_i = \left| \begin{array}{l}
{}_{[1]}c^{0:t} \wedge {}_{[1]}\alpha_i^{0:t} \wedge {}_{[2:L]}c^{0:t-1} \wedge {}_{[2:L]}\alpha_i^{0:t-1} \wedge \\
{}_{[1]}z_i^{0:t} \wedge {}_{[2:L]}z_i^{0:t-1} \wedge {}_{[1]}b^{0:t} \wedge {}_{[1]}\beta_i^{0:t} \wedge {}_{[2:L]}b^{0:t-1} \wedge {}_{[2:L]}\beta_i^{0:t-1} \wedge \\
m^{0:t-1} \wedge \lambda_i^{0:t-1} \wedge \pi_i
\end{array} \right.
$$

$$
P(S_i^t|_{[1]}\nu_e^t \wedge \pi_i)
$$

$$= P(S_i^t | \begin{bmatrix} {}_{[1]}c^{0:t} \wedge {}_{[1]}\alpha_i^{0:t} \wedge {}_{[2:L]}c^{0:t-1} \wedge {}_{[2:L]}\alpha_i^{0:t-1} \wedge \\ {}_{[1]}z_i^{0:t} \wedge {}_{[2:L]}z_i^{0:t-1} \wedge {}_{[1]}b^{0:t} \wedge {}_{[1]}\beta_i^{0:t} \wedge {}_{[2:L]}b^{0:t-1} \wedge {}_{[2:L]}\beta_i^{0:t-1} \wedge \\ m^{0:t-1} \wedge \lambda_i^{0:t-1} \wedge \pi_i \end{bmatrix})$$

$$P(S_i^t|_{[1]}\nu_e^t \wedge \pi_i)$$

$$= \left| \begin{array}{l} \frac{1}{P({}_{[1]}\nu_e^t|\pi_i)} \\ \\ \times \sum \begin{bmatrix} S_i^{0:t-1} \\ {}_{[2:L]}C^t \\ {}_{[2:L]}Z_i^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha_i^t \\ {}_{[2:L]}\beta_i^t \\ \lambda_i^t \end{bmatrix} \begin{bmatrix} \prod_{j=1}^{t} \begin{bmatrix} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\ \times P({}_{[1]}\beta_i^j|_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\ \times P({}_{[1]}\alpha_i^j|_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\ \times P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \begin{bmatrix} P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\ \times P({}_{[k]}\beta_i^j|_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^j|_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \times P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{bmatrix} \\ \times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i) \end{bmatrix} \\ \times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i) \end{bmatrix} \end{array} \right|$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_e^t|\pi_i)} \\
\times \sum \begin{bmatrix} S_i^{0:t-1} \\ {}_{[2:L]}C^t \\ {}_{[2:L]}Z_i^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha_i^t \\ {}_{[2:L]}\beta_i^t \\ \lambda_i^t \end{bmatrix}
\begin{bmatrix}
P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{L} \begin{bmatrix}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{bmatrix} \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\
\prod_{j=1}^{t-1} \begin{bmatrix}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \begin{bmatrix}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{bmatrix} \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{bmatrix} \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_e^t|\pi_i)} \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\sum \left[\begin{array}{l} {}_{[2:L]}C^t \\ {}_{[2:L]}Z_i^t \\ {}_{[2:L]}B^t \\ {}_{[2:L]}\alpha_i^t \\ {}_{[2:L]}\beta_i^t \\ M^t \\ \lambda_i^t \end{array} \right] \left[\begin{array}{l}
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right] \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right] \right|
$$

Sum over lower levels is equal to one, as shown below.

$$
\sum_{\left[\begin{array}{l} {}_{[2:L]}C^t \\ {}_{[2:L]}Z_i^t \\ {}_{[2:L]}B^t \\ {}_{[2:L]}\alpha_i^t \\ {}_{[2:L]}\beta_i^t \\ M^t \\ \lambda_i^t \end{array} \right]} \left[\begin{array}{l}
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right]
$$

$$
= \sum_{{}_{[2:L]}B^t} \left[\begin{array}{l} \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi_i) \times \sum_{{}_{[k]}\beta_i^t} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times \sum_{{}_{[k]}C^t} \left[\begin{array}{l} P({}_{[k]}C^t|\pi_i) \\ \times \sum_{{}_{[k]}\alpha_i^t} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \times \sum_{{}_{[k]}Z_i^t} P({}_{[k]}Z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array} \right] \end{array} \right] \\ \times \sum_{M^t} \left[\begin{array}{l} P(M^t|\pi_i) \\ \times \sum_{\lambda_i^t} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array} \right] \end{array} \right]
$$

Estimation is calculated employing prediction expression.

$$
\begin{array}{l} P(S_i^t|{}_{[1]}\nu_e^t \wedge \pi_i) \\ = \left| \begin{array}{l} \frac{1}{P({}_{[1]}\nu_e^t|\pi_i)} \\ \times \sum_{S_i^{0:t-1}} \left[\begin{array}{l} P(S_i^t \,\big|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\ \times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\ \times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\ \times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S_i^j \,\big|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\ \times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\ \times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\ \times P({}_{[1]}Z_i^j \,\big|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\ \times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \times P({}_{[k]}Z_i^j \,\big|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{array} \right] \\ \times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i) \end{array} \right] \\ \times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i) \end{array} \right] \end{array} \right.
$$

$$
= \left|
\begin{array}{l}
\frac{1}{P({}_{[1]}\nu_e^t|\pi_i)} \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\sum_{S_i^{0:t-1}} \left[
\begin{array}{l}
P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\prod_{j=1}^{t-1} \left[
\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[
\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array}
\right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array}
\right] \\
\times P(S_i^0 m^0 {}_{[1:L]}b^0 {}_{[1:L]}c^0 {}_{[1:L]}z_i^0 {}_{[0:L]}\beta_i^0 {}_{[1:L]}\alpha_i^0 \lambda_i^0 \pi_i)
\end{array}
\right]
\end{array}
\right|
$$

$$
= \left|
\begin{array}{l}
\frac{P(\nu_p^t|\pi_i)}{P({}_{[1]}\nu_e^t|\pi_i)} \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array}
\right|
$$

Estimation at level $2 \leq h \leq L$

$$
{}_{[h]}\nu_e^t \wedge \pi_i = \left|
\begin{array}{l}
{}_{[1:h]}c^{0:t} \wedge {}_{[1:h]}\alpha_i^{0:t} \wedge {}_{[h+1:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha_i^{0:t-1} \wedge \\
{}_{[1:h]}b^{0:t} \wedge {}_{[1:h]}\beta_i^{0:t} \wedge {}_{[h+1:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta_i^{0:t-1} \wedge \\
{}_{[1:h]}z_i^{0:t} \wedge {}_{[h+1:L]}z_i^{0:t-1} \wedge m^{0:t-1} \wedge \lambda_i^{0:t-1} \wedge \pi_i
\end{array}
\right.
$$

$$
\begin{array}{l}
P(S_i^t|{}_{[h]}\nu_e^t \wedge \pi_i) \\
\quad = P(S_i^t| \left[
\begin{array}{l}
{}_{[1:h]}c^{0:t} \wedge {}_{[1:h]}\alpha_i^{0:t} \wedge {}_{[h+1:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha_i^{0:t-1} \wedge \\
{}_{[1:h]}b^{0:t} \wedge {}_{[1:h]}\beta_i^{0:t} \wedge {}_{[h+1:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta_i^{0:t-1} \wedge \\
{}_{[1:h]}z_i^{0:t} \wedge {}_{[h+1:L]}z_i^{0:t-1} \wedge m^{0:t-1} \wedge \lambda_i^{0:t-1} \wedge \pi_i
\end{array}
\right])
\end{array}
$$

$P(S_i^t|{}_{[h]}\nu_e^t \wedge \pi_i)$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_e^t|\pi_i)} \\
\times \sum \left[\begin{array}{l} S_i^{0:t-1} \\ {}_{[h+1:L]}C^t \\ {}_{[h+1:L]}Z_i^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha_i^t \\ {}_{[h+1:L]}\beta_i^t \\ \lambda_i^t \end{array} \right]
\left[\begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_e^t|\pi_i)} \\
\times \sum \left[\begin{array}{l} S_i^{0:t-1} \\ {}_{[h+1:L]}C^t \\ {}_{[h+1:L]}Z_i^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha_i^t \\ {}_{[h+1:L]}\beta_i^t \\ \lambda_i^t \end{array} \right]
\left[\begin{array}{l}
P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|\, S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t \,|\, S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times \prod_{k=h+1}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|\, S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_e^t|\pi_i)} \\
\times \sum_{\left[\ S_i^{0:t-1}\ \right]} \left[\begin{array}{l}
P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times \sum \left[\begin{array}{l}
{}_{[h+1:L]}C^t \\
{}_{[h+1:L]}Z_i^t \\
{}_{[h+1:L]}B^t \\
M^t \\
{}_{[h+1:L]}\alpha_i^t \\
{}_{[h+1:L]}\beta_i^t \\
\lambda_i^t
\end{array} \right] \left[\begin{array}{l}
\prod_{k=h+1}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\times P({}_{[k]}Z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right] \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right]
\end{array} \right|
$$

It is possible to show that summing over variables of lower levels in the hierarchy is equal to one.

$$
\sum_{\left[\begin{array}{l} {}_{[h+1:L]}C^t \\ {}_{[h+1:L]}Z_i^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha_i^t \\ {}_{[h+1:L]}\beta_i^t \\ \lambda_i^t \end{array}\right]} \left[\begin{array}{l} \prod_{k=h+1}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}C^t|\pi_i) \\ \times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \times P({}_{[k]}Z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array}\right] \\ \times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array}\right]
$$

$$
= \sum_{{}_{[h+1:L]}B^t} \left[\begin{array}{l} \prod_{k=h+1}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi_i) \times \sum_{{}_{[k]}\beta_i^t} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}B^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times \sum_{{}_{[k]}C^t} \left[\begin{array}{l} P({}_{[k]}C^t|\pi_i) \\ \times \sum_{{}_{[k]}\alpha_i^t} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \times \sum_{{}_{[k]}Z_i^t} P({}_{[k]}Z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array}\right] \end{array}\right] \\ \times \sum_{M^t} \left[\begin{array}{l} P(M^t|\pi_i) \\ \times \sum_{\lambda_i^t} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array}\right] \end{array}\right]
$$

$$
P(S_i^t|{}_{[h]}\nu_e^t \wedge \pi_i)
$$

$$
= \left|\begin{array}{l} \frac{1}{P({}_{[h]}\nu_e^t|\pi_i)} \\ \times \sum_{\left[\ S_i^{0:t-1}\ \right]} \left[\begin{array}{l} P(S_i^t \,\big|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\ \times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\ \times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\ \times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{k=2}^{h} \left[\begin{array}{l} P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\ \times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\ \times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i) \end{array}\right] \\ \times \prod_{j=1}^{t-1} \left[\begin{array}{l} P(S_i^j \,\big|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\ \times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\ \times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\ \times P({}_{[1]}Z_i^j \,\big|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\ \times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \times P({}_{[k]}Z_i^j \,\big|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{array}\right] \\ \times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i) \end{array}\right] \\ \times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i) \end{array}\right]
$$

$$
= \left|
\begin{array}{l}
\frac{1}{P({}_{[h]}\nu_e^t|\pi_i)} \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times \sum_{\left[\ S_i^{0:t-1}\ \right]} \left[\begin{array}{l}
P(S_i^t \,\big|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,\big|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,\big|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,\big|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left|
\begin{array}{l}
\frac{P(\nu_p^t|\pi_i)}{P({}_{[h]}\nu_e^t|\pi_i)} \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array} \right|
$$

E.1.6 Motor Commands

$$
\nu_m^t \wedge \pi_i = {}_{[1:L]}c^{0:t} \wedge {}_{[1:L]}\alpha_i^{0:t} \wedge {}_{[1:L]}z_i^{0:t} \wedge {}_{[1:L]}b^{0:t} \wedge {}_{[1:L]}\beta_i^{0:t} \wedge m^{0:t-1} \wedge \lambda_i^{0:t} \wedge \pi_i
$$

$$
\begin{aligned}
&P(M^t|\nu_m^t \wedge \pi_i) \\
&\quad = P(M^t|\ {}_{[1:L]}c^{0:t} \wedge {}_{[1:L]}\alpha_i^{0:t} \wedge {}_{[1:L]}z_i^{0:t} \wedge {}_{[1:L]}b^{0:t} \wedge {}_{[1:L]}\beta_i^{0:t} \wedge m^{0:t-1} \wedge \lambda_i^{0:t} \wedge \pi_i\ )
\end{aligned}
$$

$$
\begin{aligned}
&P(M^t|\nu_m^t \wedge \pi_i) \\
&= \left| \begin{array}{l}
\frac{1}{P(\nu_m^t|\pi_i)} \\
\times \sum_{S_i^{0:t}} \left[\begin{array}{l}
P(S_i^0 m^0{}_{[1:L]} b^0{}_{[1:L]} c^0{}_{[1:L]} z_i^0{}_{[0:L]} \beta_i^0{}_{[1:L]} \alpha_i^0 \lambda_i^0 \pi_i) \\
\times \\
\prod_{j=1}^{t} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right]
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{1}{P(\nu_m^t|\pi_i)} \\
\times \sum_{S_i^{0:t}} \left[\begin{array}{l}
P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}b^t \wedge S_i^t\pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S_i^0 m^0{}_{[1:L]} b^0{}_{[1:L]} c^0{}_{[1:L]} z_i^0{}_{[0:L]} \beta_i^0{}_{[1:L]} \alpha_i^0 \lambda_i^0 \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left|\begin{array}{l}
\frac{1}{P(\nu_m^t|\pi_i)} \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array}\right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\
\times \sum_{S_i^{0:t-1}} \left[\begin{array}{l}
P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi_i) \times P({}_{[1]}C^j|\pi_i) \\
\times P({}_{[1]}\beta_i^j|_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^j|_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\times P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi_i) \times P({}_{[k]}C^j|\pi_i) \\
\times P({}_{[k]}\beta_i^j|_{[k]}B^j \wedge {}_{[k-1]}B^j \wedge S_i^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^j|_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\times P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array}\right] \\
\times P(M^j|\pi_i) \times P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array}\right] \\
\times P(S_i^0 m^0{}_{[1:L]}b^0{}_{[1:L]}c^0{}_{[1:L]}z_i^0{}_{[0:L]}\beta_i^0{}_{[1:L]}\alpha_i^0\lambda_i^0\pi_i)
\end{array}\right]
\end{array}\right| \\
&= \left|\begin{array}{l}
\frac{P(\nu_p^t|\pi_i)}{P(\nu_m^t|\pi_i)} \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array}\right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}b^t \wedge S_i^t\pi_i) \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array}\right]
\end{array}\right| \\
&= \left|\begin{array}{l}
\frac{P({}_{[L]}\nu_e^t|\pi_i)}{P(\nu_m^t|\pi_i)} \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}b^t \wedge S_i^t\pi_i) \\
\times P(S_i^t|_{[t]}\nu_e^{\wedge}\pi_i)
\end{array}\right]
\end{array}\right|
\end{aligned}
\tag{E.6}
$$

E.1.7 Prediction Loop

It is possible to express prediction in function of estimation in previous time step. Following as stated in equation E.3.

$$
\begin{aligned}
&P(S_i^t|\nu_p^t \wedge \pi_i) \\
&= \left| \begin{array}{l} \frac{P(\nu_p^{t-1}|\pi_i)}{P(\nu_p^t|\pi_i)} \\ \times \sum_{S_i^{t-1}} \left[\begin{array}{l} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\ \times P({}_{[1]}b^{t-1}|\pi_i) \times P({}_{[1]}c^{t-1}|\pi_i) \times \\ P({}_{[1]}\beta_i^{t-1}|_{[1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-2}\pi_i) \\ \times P({}_{[1]}\alpha_i^{t-1}|_{[1]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-1}\pi_i) \\ \times P({}_{[1]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[1]}c^{t-1} \wedge \pi_i) \\ \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}b^{t-1}|\pi_i) \times P({}_{[k]}c^{t-1}|\pi_i) \\ \times P({}_{[k]}\beta_i^{t-1}|_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-2}\pi_i) \\ \times P({}_{[k]}\alpha_i^{t-1}|_{[k]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1}\pi_i) \\ \times P({}_{[k]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[k]}c^{t-1} \wedge \pi_i) \end{array} \right] \\ \times P(m^{t-1}|\pi_i) \times P(\lambda_i^{t-1}|m^{t-1} \wedge m^{t-2} \wedge {}_{[L]}b^{t-1} \wedge S_i^{t-1}\pi_i) \\ \times P(S_i^{t-1}|\nu_p^{t-1} \wedge \pi_i) \end{array} \right] \end{array} \right| \\
&= \left| \begin{array}{l} \frac{P(\nu_p^{t-1}|\pi_i)}{P(\nu_p^t|\pi_i)} \\ \times \sum_{S_i^{t-1}} \left[\begin{array}{l} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\ \times P(m^{t-1}|\pi_i) \\ \times P(\lambda_i^{t-1}|m^{t-1} \wedge m^{t-2} \wedge {}_{[L]}b^{t-1} \wedge S_i^{t-1}\pi_i) \\ \times P(S_i^{t-1}|_{[L]}\nu_e^{t-1} \wedge \pi_i) \end{array} \right] \end{array} \right|
\end{aligned}
\tag{E.7}
$$

E.1.8 Attention and Behaviour at level h, with $2 \leq h \leq L$

Lets look at expression for attention expression at a general level h, with $2 \leq h \leq L$:

$$P({}_{[h]}C^t \wedge {}_{[h]}B^t|_{[h]}\nu_a^t \wedge \pi_i)$$

$$
= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi_i)}{P({}_{[h]}\nu_a^t|\pi_i)} \\ \times \sum_{S_i^t} \left[\begin{array}{l} P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\ \times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\ \times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{k=2}^{h-1} \left[\begin{array}{l} P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\ \times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\ \times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i) \end{array} \right] \\ \times P({}_{[h]}B^t|\pi_i) \times P({}_{[h]}C^t|\pi_i) \\ \times P({}_{[h]}\beta_i^t|{}_{[h]}B^t \wedge {}_{[h-1]}b^t \wedge S_i^t \wedge {}_{[h]}b^{t-1}\pi_i) \\ \times P({}_{[h]}\alpha_i^t|{}_{[h]}C^t \wedge S_i^t \wedge {}_{[h]}B^t \wedge {}_{[h-1]}b^t\pi_i) \\ \times P(S_i^t|\nu_p^t \wedge \pi_i) \end{array} \right] \end{array} \right|
$$

It is possible to see that it function of prediction, but it is possible to go farther and write it in relation to estimation at previous level $h-1$. Estimation is as stated in equation E.10, and we find:

$$
\begin{aligned}
& P({}_{[h]}C^t \wedge {}_{[h]}B^t|{}_{[h]}\nu_a^t \wedge \pi_i) \\
& = \left| \begin{array}{l} \frac{P(\nu_p^t|\pi_i)}{P({}_{[h]}\nu_a^t|\pi_i)} \times \frac{P({}_{[h-1]}\nu_e^t|\pi_i)}{P(\nu_p^t|\pi_i)} \\ \times \sum_{S_i^t} \left[\begin{array}{l} P({}_{[h]}B^t|\pi_i) \times P({}_{[h]}C^t|\pi_i) \\ \times P({}_{[h]}\beta_i^t|{}_{[h]}B^t \wedge {}_{[h-1]}b^t \wedge S_i^t \wedge {}_{[h]}b^{t-1}\pi_i) \\ \times P({}_{[h]}\alpha_i^t|{}_{[h]}C^t \wedge S_i^t \wedge {}_{[h]}B^t \wedge {}_{[h-1]}b^t\pi_i) \\ \times P(S_i^t|{}_{[h-1]}\nu_e^t \wedge \pi_i) \end{array} \right] \end{array} \right|
\end{aligned}
\tag{E.8}
$$

The same principle applies for behaviour expression at a general level h, with $2 \leq h \leq L$:

$$
\begin{aligned}
& P({}_{[h]}B^t|{}_{[h]}\nu_b^t \wedge \pi_i) \\
& = \left| \begin{array}{l} \frac{P(\nu_p^t|\pi_i)}{P({}_{[h]}\nu_b^t|\pi_i)} \\ \times \sum_{S_i^t} \left[\begin{array}{l} P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \times P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\ \times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{k=2}^{h} \left[\begin{array}{l} P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\ \times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\ \times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i) \end{array} \right] \\ \times P(S_i^t|\nu_p^t \wedge \pi_i) \end{array} \right] \end{array} \right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi_i)}{P({}_{[h]}\nu_b^t|\pi_i)} \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h-1} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P({}_{[h]}b^t|\pi_i) \times P({}_{[h]}\beta_i^t|_{[h]}b^t \wedge {}_{[h-1]}b^t \wedge S_i^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\times P({}_{[h]}c^t|\pi_i) \times P({}_{[h]}\alpha_i^t|_{[h]}c^t \wedge S_i^t \wedge {}_{[h]}b^t \wedge {}_{[h-1]}b^t\pi_i) \\
\times P({}_{[h]}z_i^t\,|S_i^t \wedge {}_{[h]}c^t \wedge \pi_i) \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi_i)}{P({}_{[h]}\nu_b^t|\pi_i)} \times \frac{P({}_{[h-1]}\nu_e^t|\pi_i)}{P(\nu_p^t|\pi_i)} \\
\times \sum_{S_i^t} \left[\begin{array}{l}
P({}_{[h]}B^t|\pi_i) \times P({}_{[h]}\beta_i^t|_{[h]}B^t \wedge {}_{[h-1]}b^t \wedge S_i^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\times P({}_{[h]}c^t|\pi_i) \times P({}_{[h]}\alpha_i^t|_{[h]}c^t \wedge S_i^t \wedge {}_{[h]}B^t \wedge {}_{[h-1]}b^t\pi_i) \\
\times P({}_{[h]}z_i^t\,|S_i^t \wedge {}_{[h]}c^t \wedge \pi_i) \\
\times P(S_i^t|_{[h-1]}\nu_e^t \wedge \pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
\tag{E.9}
$$

Last, it is also possible to write estimation at level h in relation to estimation at level $h-1$.

$$
\begin{aligned}
&P(S_i^t|_{[h]}\nu_e^t \wedge \pi_i) \\
&= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi_i)}{P({}_{[h]}\nu_e^t|\pi_i)} \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array} \right|
\end{aligned}
$$

$$
\begin{array}{rl}
= & \left|\begin{array}{l}
\frac{P(\nu_p^t|\pi_i)}{P({}_{[h]}\nu_e^t|\pi_i)} \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}c^t|\pi_i) \\
\times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h-1}\left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge {}_{[k-1]}b^t \wedge S_i^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array}\right] \\
\times P({}_{[h]}b^t|\pi_i) \times P({}_{[h]}\beta_i^t|{}_{[h]}b^t \wedge {}_{[h-1]}b^t \wedge S_i^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\times P({}_{[h]}c^t|\pi_i) \times P({}_{[h]}\alpha_i^t|{}_{[h]}c^t \wedge S_i^t \wedge {}_{[h]}b^t \wedge {}_{[h-1]}b^t\pi_i) \\
\times P({}_{[h]}z_i^t\,|S_i^t \wedge {}_{[h]}c^t \wedge \pi_i) \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array}\right| \\
= & \left|\begin{array}{l}
\frac{P(\nu_p^t|\pi_i)}{P({}_{[h]}\nu_e^t|\pi_i)} \times \frac{P({}_{[h-1]}\nu_e^t|\pi_i)}{P(\nu_p^t|\pi_i)} \\
\times P({}_{[h]}b^t|\pi_i) \times P({}_{[h]}\beta_i^t|{}_{[h]}b^t \wedge {}_{[h-1]}b^t \wedge S_i^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\times P({}_{[h]}c^t|\pi_i) \times P({}_{[h]}\alpha_i^t|{}_{[h]}c^t \wedge S_i^t \wedge {}_{[h]}b^t \wedge {}_{[h-1]}b^t\pi_i) \\
\times P({}_{[h]}z_i^t\,|S_i^t \wedge {}_{[h]}c^t \wedge \pi_i) \\
\times P(S_i^t|{}_{[h-1]}\nu_e^t \wedge \pi_i)
\end{array}\right|
\end{array}
\tag{E.10}
$$

E.2 Global Filter

E.2.1 Joint Distribution

$$
\begin{array}{rl}
 & P(S^{0:t} \wedge {}_{[1:L]}C^{0:t} \wedge {}_{[1:L]}Z^{0:t} \wedge {}_{[1:L]}B^{0:t} \wedge M^{0:t} \wedge \lambda^{0:t} \wedge {}_{[1:L]}\beta^{0:t} \wedge {}_{[1:L]}\alpha^{0:t}\,|\pi) \\
= & \left|\begin{array}{l}
\prod_{j=1}^{t}\left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,\big|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,\big|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L}\left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,\big|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array}\right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array}\right] \\
\times P(S^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z^0{}_{[1:L]}\beta^0{}_{[1:L]}\alpha^0\lambda^0|\pi)
\end{array}\right|
\end{array}
$$

E.2.2 Prediction

$$
\nu_p^t = \; {}_{[1:L]}c^{0:t-1} \wedge {}_{[1:L]}\alpha^{0:t-1} \wedge {}_{[1:L]}z^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1:L]}\beta^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1}
$$

$$
\begin{aligned}
&P(S^t|\nu_p^t \wedge \pi) \\
&\quad = P\Big(S^t \Big|_{[1:L]}c^{0:t-1} \wedge {}_{[1:L]}\alpha^{0:t-1} \wedge {}_{[1:L]}z^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1:L]}\beta^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi\Big)
\end{aligned}
$$

$$
\begin{aligned}
&P(S^t|\nu_p^t \wedge \pi) \\
&= \left| \begin{array}{l}
\frac{1}{P(\nu_p^t|\pi)} \\
\times \sum \begin{bmatrix} S^{0:t-1} \\ {}_{[1:L]}C^t \\ {}_{[1:L]}Z^t \\ {}_{[1:L]}B^t \\ M^t \\ {}_{[1:L]}\alpha^t \\ {}_{[1:L]}\beta^t \\ \lambda^t \end{bmatrix} \left[\begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P\Big(S_i^j \Big| S_i^{j-1} \wedge M^{j-1} \wedge \pi_i\Big) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P\Big({}_{[1]}Z_i^j \Big| S_i^j \wedge {}_{[1]}C^j \wedge \pi_i\Big) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P\Big({}_{[k]}Z_i^j \Big| S_i^j \wedge {}_{[k]}C^j \wedge \pi_i\Big)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi)
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^t|\pi)} \\
\times \sum \left[\begin{array}{l} S^{0:t-1} \\ {}_{[1:L]}C^t \\ {}_{[1:L]}Z^t \\ {}_{[1:L]}B^t \\ M^t \\ {}_{[1:L]}\alpha^t \\ {}_{[1:L]}\beta^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^t \,|S_i^t \wedge {}_{[1]}C^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi)
\end{array} \right] \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^t|\pi)} \times \\
\sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \times \\
\sum \left[\begin{array}{l} {}_{[1:L]}C^t \\ {}_{[1:L]}Z^t \\ {}_{[1:L]}B^t \\ M^t \\ {}_{[1:L]}\alpha^t \\ {}_{[1:L]}\beta^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l}
P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^t \,|S_i^t \wedge {}_{[1]}C^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right] \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi)
\end{array} \right] \end{array} \right|
$$

It is possible to show that the sum over variables at time t sum to one.

$$
\sum \left[\begin{array}{l} {}_{[1:L]}C^t \\ {}_{[1:L]}Z^t \\ {}_{[1:L]}B^t \\ M^t \\ {}_{[1:L]}\alpha^t \\ {}_{[1:L]}\beta^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l}
P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^t \,|S_i^t \wedge {}_{[1]}C^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right]
$$

$$
= \sum_{{}_{[1:L]}B^t} \left[\begin{array}{l}
P({}_{[1]}B^t|\pi) \times \prod_{i=1}^{N_i} \sum_{{}_{[1]}\beta_i^t} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\sum_{{}_{[1]}C^t} \left[\begin{array}{l} P({}_{[1]}C^t|\pi) \\ \times \prod_{i=1}^{N_i} \sum_{{}_{[1]}\alpha_i^t} P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} \sum_{{}_{[1]}Z_i^t} P({}_{[1]}Z_i^t|S_i^t \wedge {}_{[1]}C^t \wedge \pi_i) \end{array} \right] \\
\prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi) \times \prod_{i=1}^{N_i} \sum_{{}_{[k]}\beta_i^t} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\sum_{{}_{[k]}C^t} \left[\begin{array}{l} P({}_{[k]}C^t|\pi) \\ \prod_{i=1}^{N_i} \sum_{{}_{[k]}\alpha_i^t} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} \sum_{{}_{[k]}Z_i^t} P({}_{[k]}Z_i^t|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array} \right] \end{array} \right] \\
\times \sum_{M^t} \left[\begin{array}{l} P(M^t|\pi) \\ \times \prod_{i=1}^{N_i} \sum_{\lambda_i^t} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array} \right]
\end{array} \right]
$$

And simplified expression for prediction is:

$$
\begin{array}{l}
P(S^t|\nu_p^t \wedge \pi) \\
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^t|\pi)} \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z^0{}_{[1:L]}\beta^0{}_{[1:L]}\alpha^0\lambda^0|\pi)
\end{array} \right]
\end{array} \right|
\end{array}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^t|\pi)} \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\prod_{i=1}^{N_i} P(S_i^{t-1} \,|\, S_i^{t-2} \wedge m^{t-2} \wedge \pi_i) \\
\times P({}_{[1]}b^{t-1}|\pi) \times P({}_{[1]}c^{t-1}|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^{t-1}|{}_{[1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^{t-1}|{}_{[1]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[1]}c^{t-1} \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^{t-1}|\pi) \times P({}_{[k]}c^{t-1}|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^{t-1}|{}_{[k]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[k-1]}b^{t-1} \wedge {}_{[k]}b^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^{t-1}|{}_{[k]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[k]}c^{t-1} \wedge \pi_i)
\end{array} \right] \\
\times P(m^{t-1}|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^{t-1}|m^{t-1} \wedge m^{t-2} \wedge {}_{[L]}b^{t-1} \wedge S_i^{t-1}\pi_i) \\
\prod_{j=1}^{t-2} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0{}_{[1:L]} B^0{}_{[1:L]} C^0{}_{[1:L]} Z^0{}_{[1:L]} \beta^0{}_{[1:L]} \alpha^0 \lambda^0|\pi)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^t|\pi)} \\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^{t-1}|\pi) \times P({}_{[1]}c^{t-1}|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^{t-1}|{}_{[1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^{t-1}|{}_{[1]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[1]}c^{t-1} \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^{t-1}|\pi) \times P({}_{[k]}c^{t-1}|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^{t-1}|{}_{[k]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[k-1]}b^{t-1} \wedge {}_{[k]}b^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^{t-1}|{}_{[k]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[k]}c^{t-1} \wedge \pi_i)
\end{array} \right] \\
\times P(m^{t-1}|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^{t-1}|m^{t-1} \wedge m^{t-2} \wedge {}_{[L]}b^{t-1} \wedge S_i^{t-1}\pi_i) \\
\times \sum_{S^{0:t-2}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^{t-1} \,|\, S_i^{t-2} \wedge m^{t-2} \wedge \pi_i) \\
\prod_{j=1}^{t-2} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z^0{}_{[1:L]}\beta^0{}_{[1:L]}\alpha^0\lambda^0|\pi)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^{t-1}|\pi)}{P(\nu_p^t|\pi)} \\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^{t-1}|\pi) \times P({}_{[1]}c^{t-1}|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^{t-1}|{}_{[1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^{t-1}|{}_{[1]}C^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[1]}c^{t-1} \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^{t-1}|\pi) \times P({}_{[k]}c^{t-1}|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^{t-1}|{}_{[k]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[k-1]}b^{t-1} \wedge {}_{[k]}b^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^{t-1}|{}_{[k]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^{t-1} \,|\, S_i^{t-1} \wedge {}_{[k]}c^{t-1} \wedge \pi_i)
\end{array} \right] \\
\times P(m^{t-1}|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^{t-1}|m^{t-1} \wedge m^{t-2} \wedge {}_{[L]}b^{t-1} \wedge S_i^{t-1}\pi_i) \\
\times P(S^{t-1}|\nu_p^{t-1} \wedge \pi)
\end{array} \right]
\end{array} \right|
$$

E.2.3 Attention Selection

Attention Selection at level 1

$$
{}_{[1]}\nu_a^t = {}_{[1:L]}c^{0:t-1} \wedge {}_{[1:L]}\alpha^{0:t} \wedge {}_{[1:L]}z^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1:L]}\beta^{0:t} \wedge m^{0:t-1} \wedge \lambda^{0:t-1}
$$

$$
\begin{aligned}
&P({}_{[1]}C^t \wedge {}_{[1]}B^t | {}_{[1]}\nu_a^t \wedge \pi) \\
&\quad = P({}_{[1]}C^t \wedge {}_{[1]}B^t | {}_{[1:L]}c^{0:t-1} \wedge {}_{[1:L]}\alpha^{0:t} \wedge {}_{[1:L]}z^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1:L]}\beta^{0:t} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi)
\end{aligned}
$$

$$
\begin{aligned}
&P({}_{[1]}C^t \wedge {}_{[1]}B^t | {}_{[1]}\nu_a^t \wedge \pi) \\
&= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_a^t | \pi)} \\
\times \sum \left[\begin{array}{l} S^{0:t} \\ {}_{[2:L]}C^t \\ {}_{[1:L]}Z^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha^t \\ {}_{[2:L]}\beta^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l}
\prod_{j=1}^{t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j | \pi) \times P({}_{[1]}C^j | \pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j | {}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1} \pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j | {}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j \pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j | \pi) \times P({}_{[k]}C^j | \pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j | {}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1} \pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j | {}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j \pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j | \pi) \prod_{i=1}^{N_i} P(\lambda_i^j | M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j \pi_i)
\end{array} \right] \\
\times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0 | \pi)
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_a^t|\pi)} \\
\times \sum \left[\begin{array}{l} S^{0:t} \\ {}_{[2:L]}C^t \\ {}_{[1:L]}Z^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha^t \\ {}_{[2:L]}\beta^t \\ \lambda^t \end{array} \right]
\left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^t \,|S_i^t \wedge {}_{[1]}C^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_a^t|\pi)} \\
\times \sum_{S^{0:t}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}B^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\times \sum_{{}_{[1]}Z^t} \left[\prod_{i=1}^{N_i} P({}_{[1]}Z_i^t \,|S_i^t \wedge {}_{[1]}C^t \wedge \pi_i) \right] \\
\times \sum \left[\begin{array}{l} {}_{[2:L]}C^t \\ {}_{[2:L]}Z^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha^t \\ {}_{[2:L]}\beta^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l}
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right] \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi)
\end{array} \right] \right|
$$

Sum over lower levels in the hierarchy is equal to one, as shown below.

$$
\sum_{\left[\begin{array}{l} {}_{[2:L]}C^t \\ {}_{[2:L]}Z^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha^t \\ {}_{[2:L]}\beta^t \\ \lambda^t \end{array} \right]} \left[\begin{array}{l}
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right]
$$

$$
= \sum_{[2:L]B^t} \left[\begin{array}{l} \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi) \\ \prod_{i=1}^{N_i} \sum_{{}_{[k]}\beta_i^t} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times \sum_{{}_{[k]}C^t} \left[\begin{array}{l} P({}_{[k]}C^t|\pi) \\ \prod_{i=1}^{N_i} \sum_{{}_{[k]}\alpha_i^t} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} \sum_{{}_{[k]}Z_i^t} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array} \right] \end{array} \right] \\ \times \sum_{M^t} \left[\begin{array}{l} P(M^t|\pi) \\ \times \prod_{i=1}^{N_i} \sum_{\lambda_i^t} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array} \right] \end{array} \right]
$$

$$
P({}_{[1]}C^t \wedge {}_{[1]}B^t|{}_{[1]}\nu_a^t \wedge \pi)
$$

$$
= \left| \begin{array}{l} \frac{1}{P({}_{[1]}\nu_a^t|\pi)} \\ \times \sum_{S^t} \left[\begin{array}{l} P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \times \sum_{S^{0:t-1}} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^t \,\big|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\ \prod_{j=1}^{t-1} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^j \,\big|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,\big|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,\big|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{array} \right] \\ \times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i) \end{array} \right] \\ \times P(S^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z^0{}_{[1:L]}\beta^0{}_{[1:L]}\alpha^0\lambda^0|\pi) \end{array} \right] \end{array} \right]
$$

$$
= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_a^t|\pi)} \\ \times \sum_{S^t} \left[\begin{array}{l} P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \times P(S^t|\nu_p^t \wedge \pi) \end{array} \right] \end{array} \right.
$$

Attention Selection at level h**, with** $2 \leq h \leq L$

$$
{}_{[h]}\nu_a^t = \left| \begin{array}{l} {}_{[1:h-1]}c^{0:t} \wedge {}_{[1:h]}\alpha^{0:t} \wedge {}_{[h:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha^{0:t-1} \wedge \\ {}_{[1:h-1]}b^{0:t} \wedge {}_{[1:h]}\beta^{0:t} \wedge {}_{[h:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta^{0:t-1} \wedge \\ {}_{[1:h-1]}z^{0:t} \wedge {}_{[h:L]}z^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \end{array} \right.
$$

$$
P({}_{[h]}C^t \wedge {}_{[h]}B^t|{}_{[h]}\nu_a^t \wedge \pi)
$$

$$= P\Big({}_{[h]}C^t \wedge {}_{[h]}B^t \Big| \begin{bmatrix} {}_{[1:h-1]}c^{0:t} \wedge {}_{[1:h]}\alpha^{0:t} \wedge {}_{[h:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha^{0:t-1} \wedge \\ {}_{[1:h-1]}b^{0:t} \wedge {}_{[1:h]}\beta^{0:t} \wedge {}_{[h:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta^{0:t-1} \wedge \\ {}_{[1:h-1]}z^{0:t} \wedge {}_{[h:L]}z^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi \end{bmatrix}\Big)$$

$$P\big({}_{[h]}C^t \wedge {}_{[h]}B^t \big| {}_{[h]}\nu_a^t \wedge \pi\big)$$

$$= \left| \begin{array}{l} \frac{1}{P({}_{[h]}\nu_a^t|\pi)} \\ \times \sum \begin{bmatrix} S^{0:t} \\ {}_{[h+1:L]}C^t \\ {}_{[h:L]}Z^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha^t \\ {}_{[h+1:L]}\beta^t \\ \lambda^t \end{bmatrix} \begin{bmatrix} \prod_{j=1}^{t} \begin{bmatrix} \prod_{i=1}^{N_i} P\big(S_i^j \big| S_i^{j-1} \wedge M^{j-1} \wedge \pi_i\big) \\ \times P\big({}_{[1]}B^j|\pi\big) \times P\big({}_{[1]}C^j|\pi\big) \\ \prod_{i=1}^{N_i} P\big({}_{[1]}\beta_i^j|_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i\big) \\ \prod_{i=1}^{N_i} P\big({}_{[1]}\alpha_i^j|_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i\big) \\ \prod_{i=1}^{N_i} P\big({}_{[1]}Z_i^j \big| S_i^j \wedge {}_{[1]}C^j \wedge \pi_i\big) \\ \prod_{k=2}^{L} \begin{bmatrix} P\big({}_{[k]}B^j|\pi\big) \times P\big({}_{[k]}C^j|\pi\big) \\ \prod_{i=1}^{N_i} P\big({}_{[k]}\beta_i^j|_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i\big) \\ \prod_{i=1}^{N_i} P\big({}_{[k]}\alpha_i^j|_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i\big) \\ \prod_{i=1}^{N_i} P\big({}_{[k]}Z_i^j \big| S_i^j \wedge {}_{[k]}C^j \wedge \pi_i\big) \end{bmatrix} \\ \times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i) \end{bmatrix} \\ \times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi) \end{bmatrix} \end{array} \right|$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_a^t|\pi)} \\
\times \sum \left[\begin{array}{l} S^{0:t} \\ {}_{[h+1:L]}C^t \\ {}_{[h:L]}Z^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha^t \\ {}_{[h+1:L]}\beta^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^t \,|S_i^t \wedge {}_{[1]}C^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_a^t|\pi)} \\
\times \sum_{S^{0:t}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h-1} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi) \times P({}_{[k]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P({}_{[h]}B^t|\pi) P({}_{[h]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[h]}\beta_i^t|{}_{[h]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[h]}\alpha_i^t|{}_{[h]}C^t \wedge S_i^t \wedge {}_{[h]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times \sum_{{}_{[h]}Z^t} \left[\prod_{i=1}^{N_i} P({}_{[h]}Z_i^t \,|S_i^t \wedge {}_{[h]}C^t \wedge \pi_i) \right] \\
\times \sum \left[\begin{array}{l}
{}_{[h+1:L]}C^t \\
{}_{[h+1:L]}Z^t \\
{}_{[h+1:L]}B^t \\
M^t \\
{}_{[h+1:L]}\alpha^t \\
{}_{[h+1:L]}\beta^t \\
\lambda^t
\end{array} \right] \left[\begin{array}{l}
\prod_{k=h+1}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right] \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z^0{}_{[1:L]}\beta^0{}_{[1:L]}\alpha^0\lambda^0|\pi)
\end{array} \right]
\end{array} \right|
$$

$$\sum_{\left[\begin{array}{l} {}_{[h+1:L]}C^t \\ {}_{[h+1:L]}Z^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha^t \\ {}_{[h+1:L]}\beta^t \\ \lambda^t \end{array}\right]} \left[\begin{array}{l} \prod_{k=h+1}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}Z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array}\right] \\ \times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array}\right]$$

$$= \sum_{{}_{[h+1:L]}B^t} \left[\begin{array}{l} \prod_{k=h+1}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi) \times \prod_{i=1}^{N_i} \sum_{btLkit} P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \sum_{{}_{[k]}C^t} \left[\begin{array}{l} P({}_{[k]}C^t|\pi) \\ \prod_{i=1}^{N_i} \sum_{{}_{[k]}\alpha_i^t} P({}_{[k]}\alpha_i^t|_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} \sum_{{}_{[k]}Z_i^t} P({}_{[k]}Z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array}\right] \end{array}\right] \\ \times \sum_{M^t} \left[\begin{array}{l} P(M^t|\pi) \\ \times \prod_{i=1}^{N_i} \sum_{\lambda_i^t} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array}\right] \end{array}\right]$$

$$P({}_{[h]}C^t \wedge {}_{[h]}B^t|_{[h]}\nu_a^t \wedge \pi)$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_a^t|\pi)} \\
\times \sum_{S^{0:t}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h-1} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi) \times P({}_{[k]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P({}_{[h]}B^t|\pi) \times P({}_{[h]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[h]}\beta_i^t|{}_{[h]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[h]}\alpha_i^t|{}_{[h]}C^t \wedge S_i^t \wedge {}_{[h]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z^0{}_{[1:L]}\beta^0{}_{[1:L]}\alpha^0\lambda^0|\pi)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_a^t|\pi)} \\
\times \sum_{S^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h-1} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi) \times P({}_{[k]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P({}_{[h]}B^t|\pi) \times P({}_{[h]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[h]}\beta_i^t|{}_{[h]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[h]}\alpha_i^t|{}_{[h]}C^t \wedge S_i^t \wedge {}_{[h]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0{}_{[1:L]} B^0{}_{[1:L]} C^0{}_{[1:L]} Z^0{}_{[1:L]} \beta^0{}_{[1:L]} \alpha^0 \lambda^0|\pi)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_a^t|\pi)} \\
\times \sum_{S^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h-1} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi) \times P({}_{[k]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P({}_{[h]}B^t|\pi) \times P({}_{[h]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[h]}\beta_i^t|{}_{[h]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[h]}\alpha_i^t|{}_{[h]}C^t \wedge S_i^t \wedge {}_{[h]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P(S^t|\nu_p^t \wedge \pi)
\end{array} \right]
\end{array} \right|
$$

E.2.4 Behavior Selection

Action at level 1

$$
{}_{[1]}\nu_b^t = \left[\begin{array}{l} {}_{[1]}c^{0:t} \wedge {}_{[1]}\alpha^{0:t} \wedge {}_{[2:L]}c^{0:t-1} \wedge {}_{[2:L]}\alpha^{0:t-1} \wedge \\ {}_{[1]}z^{0:t} \wedge {}_{[2:L]}z^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1]}\beta^{0:t} \wedge {}_{[2:L]}\beta^{0:t-1} \wedge \\ m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi \end{array} \right]
$$

$$
\begin{aligned}
&P({}_{[1]}B^t | {}_{[1]}\nu_b^t \wedge \pi) \\
&\quad = P({}_{[1]}B^t | \left[\begin{array}{l} {}_{[1]}c^{0:t} \wedge {}_{[1]}\alpha^{0:t} \wedge {}_{[2:L]}c^{0:t-1} \wedge {}_{[2:L]}\alpha^{0:t-1} \wedge \\ {}_{[1]}z^{0:t} \wedge {}_{[2:L]}z^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1]}\beta^{0:t} \wedge {}_{[2:L]}\beta^{0:t-1} \wedge \\ m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi \end{array} \right])
\end{aligned}
$$

$$
\begin{aligned}
&P({}_{[1]}B^t | {}_{[1]}\nu_b^t \wedge \pi) \\
&= \left| \begin{array}{l} \frac{1}{P({}_{[1]}\nu_b^t | \pi)} \\ \times \sum \left[\begin{array}{l} S^{0:t} \\ {}_{[2:L]}C^t \\ {}_{[2:L]}Z^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha^t \\ {}_{[2:L]}\beta^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j | \pi) \times P({}_{[1]}C^j | \pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j | {}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1} \pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j | {}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j \pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^j | \pi) \times P({}_{[k]}C^j | \pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j | {}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1} \pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j | {}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j \pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{array} \right] \\ \times P(M^j | \pi) \prod_{i=1}^{N_i} P(\lambda_i^j | M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j \pi_i) \end{array} \right] \\ \times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0 | \pi) \end{array} \right] \end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_b^t|\pi)} \\
\times \sum \left[\begin{array}{l} S^{0:t} \\ {}_{[2:L]}C^t \\ {}_{[2:L]}Z^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha^t \\ {}_{[2:L]}\beta^t \\ \lambda^t \end{array} \right]
\left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}B^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_b^t|\pi)} \\
\times \sum_{S^{0:t}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}B^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\times \sum \left[\begin{array}{l} {}_{[2:L]}C^t \\ {}_{[2:L]}Z^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha^t \\ {}_{[2:L]}\beta^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l}
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi)
\end{array} \right] \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi)
\end{array} \right] \right.
\tag{E.11}
$$

Sum over lower levels in the hierarchy is equal to one.

$$
\sum_{\left[\begin{array}{l} {}_{[2:L]}C^t \\ {}_{[2:L]}Z^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha^t \\ {}_{[2:L]}\beta^t \\ \lambda^t \end{array} \right]} \left[\begin{array}{l}
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi)
\end{array} \right]
$$

$$
= \sum_{{}_{[2:L]}B^t} \left[\begin{array}{l}
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times \prod_{i=1}^{N_i} \sum_{{}_{[k]}\beta_i^t} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times \sum_{{}_{[k]}C^t} \left[\begin{array}{l}
P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} \sum_{{}_{[k]}\alpha_i^t} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} \sum_{{}_{[k]}Z_i^t} P({}_{[k]}Z_i^t|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right]
\end{array} \right] \\
\times \sum_{M^t} \left[\begin{array}{l}
P(M^t|\pi) \\
\times \prod_{i=1}^{N_i} \sum_{\lambda_i^t} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi)
\end{array} \right]
\end{array} \right]
$$

And behaviour expression can be written in function of prediction expression:

$$
\begin{array}{l}
P({}_{[1]}B^t|{}_{[1]}\nu_b^t \wedge \pi) \\
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_b^t|\pi)} \\
\times \sum_{S^t} \left[\begin{array}{l}
P({}_{[1]}B^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z^0{}_{[1:L]}\beta^0{}_{[1:L]}\alpha^0\lambda^0|\pi)
\end{array} \right]
\end{array} \right. \\
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_b^t|\pi)} \\
\times \sum_{S^t} \left[\begin{array}{l}
P({}_{[1]}B^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\times P(S^t|\nu_p^t \wedge \pi)
\end{array} \right]
\end{array} \right.
\end{array}
$$

Action at level $2 \leq h \leq L$

$$
{}_{[h]}\nu_b^t = \left[\begin{array}{l}
{}_{[1:h]}c^{0:t} \wedge {}_{[1:h]}\alpha^{0:t} \wedge {}_{[h+1:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha^{0:t-1} \wedge \\
{}_{[1:h-1]}b^{0:t} \wedge {}_{[1:h]}\beta^{0:t} \wedge {}_{[h+1:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta^{0:t-1} \wedge \\
{}_{[1:h]}z^{0:t} \wedge {}_{[h+1:L]}z^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1}
\end{array} \right]
$$

$$
\begin{aligned}
&P({}_{[h]}B^t|{}_{[h]}\nu_b^t \wedge \pi) \\
&\quad = P({}_{[h]}B^t| \begin{bmatrix} {}_{[1:h]}c^{0:t} \wedge {}_{[1:h]}\alpha^{0:t} \wedge {}_{[h+1:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha^{0:t-1} \wedge \\ {}_{[1:h-1]}b^{0:t} \wedge {}_{[1:h]}\beta^{0:t} \wedge {}_{[h+1:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta^{0:t-1} \wedge \\ {}_{[1:h]}z^{0:t} \wedge {}_{[h+1:L]}z^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi \end{bmatrix})
\end{aligned}
$$

$$
\begin{aligned}
&P({}_{[h]}B^t{}_{[h]}\nu_b^t \wedge \pi) \\
&= \left| \begin{array}{l} \frac{1}{P({}_{[h]}\nu_b^t|\pi)} \\ \times \sum \begin{bmatrix} S^{0:t} \\ {}_{[h+1:L]}C^t \\ {}_{[h+1:L]}Z^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha^t \\ {}_{[h+1:L]}\beta^t \\ \lambda^t \end{bmatrix} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\ \times P({}_{[1]}B^t|\pi) \times P({}_{[1]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{k=2}^{L} \begin{bmatrix} P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{bmatrix} \\ \times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\ \prod_{j=1}^{t-1} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \begin{bmatrix} P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{bmatrix} \\ \times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i) \end{bmatrix} \\ \times P(S^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z^0{}_{[1:L]}\beta^0{}_{[1:L]}\alpha^0\lambda^0|\pi) \end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_b^t|\pi)} \\
\times \sum_{S^{0:t}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times \sum \left[\begin{array}{l}
{}_{[h+1:L]}C^t \\
{}_{[h+1:L]}Z^t \\
{}_{[h+1:L]}B^t \\
M^t \\
{}_{[h+1:L]}\alpha^t \\
{}_{[h+1:L]}\beta^t \\
\lambda^t
\end{array} \right] \left[\begin{array}{l}
\prod_{k=h+1}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right] \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z^0{}_{[1:L]}\beta^0{}_{[1:L]}\alpha^0\lambda^0|\pi)
\end{array} \right]
\end{array} \right|
$$

Summing over variables of lower level in the hierarchy results in one.

$$\sum_{\begin{bmatrix} {}_{[h+1:L]}C^t \\ {}_{[h+1:L]}Z^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha^t \\ {}_{[h+1:L]}\beta^t \\ \lambda^t \end{bmatrix}} \begin{bmatrix} \prod_{k=h+1}^{L} \begin{bmatrix} P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}Z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{bmatrix} \\ \times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{bmatrix}$$

$$= \sum_{{}_{[h+1:L]}B^t} \begin{bmatrix} \prod_{k=h+1}^{L} \begin{bmatrix} P({}_{[k]}B^t|\pi) \times \prod_{i=1}^{N_i} \sum_{{}_{[k]}\beta_i^t} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \sum_{{}_{[k]}C^t} \begin{bmatrix} P({}_{[k]}C^t|\pi) \\ \prod_{i=1}^{N_i} \sum_{{}_{[k]}\alpha_i^t} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} \sum_{{}_{[k]}Z_i^t} P({}_{[k]}Z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{bmatrix} \end{bmatrix} \\ \times \sum_{M^t} \begin{bmatrix} P(M^t|\pi) \\ \times \prod_{i=1}^{N_i} \sum_{\lambda_i^t} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{bmatrix} \end{bmatrix}$$

And behaviour question can be written in relation to prediction expression.

$P({}_{[h]}B^t{}_{[h]}\nu_b^t \wedge \pi)$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_b^t|\pi)} \\
\times \sum_{S^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,\big|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,\big|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,\big|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,\big|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z^0{}_{[1:L]}\beta^0{}_{[1:L]}\alpha^0\lambda^0|\pi)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_b^t|\pi)} \\
\times \sum_{S^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}B^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P(S^t|\nu_p^t \wedge \pi)
\end{array} \right]
\end{array} \right|
$$

E.2.5 Estimation

Estimation at level 1

$$
{}_{[1]}\nu_e^t = \left[\begin{array}{l}
{}_{[1]}c^{0:t} \wedge {}_{[1]}\alpha^{0:t} \wedge {}_{[2:L]}c^{0:t-1} \wedge {}_{[2:L]}\alpha^{0:t-1} \wedge \\
{}_{[1]}z^{0:t} \wedge {}_{[2:L]}z^{0:t-1} \wedge {}_{[1]}b^{0:t} \wedge {}_{[1]}\beta^{0:t} \wedge \\
{}_{[2:L]}b^{0:t-1} \wedge {}_{[2:L]}\beta^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi
\end{array} \right]
$$

$$
\begin{aligned}
&P(S^t|_{[1]}\nu_e^t \wedge \pi) \\
&\quad = P(S^t| \begin{bmatrix} {}_{[1]}c^{0:t} \wedge {}_{[1]}\alpha^{0:t} \wedge {}_{[2:L]}c^{0:t-1} \wedge {}_{[2:L]}\alpha^{0:t-1} \wedge \\ {}_{[1]}z^{0:t} \wedge {}_{[2:L]}z^{0:t-1} \wedge {}_{[1]}b^{0:t} \wedge {}_{[1]}\beta^{0:t} \wedge \\ {}_{[2:L]}b^{0:t-1} \wedge {}_{[2:L]}\beta^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi \end{bmatrix})
\end{aligned}
$$

$$
\begin{aligned}
&P(S^t|_{[1]}\nu_e^t \wedge \pi) \\
&= \left| \begin{array}{l} \frac{1}{P({}_{[1]}\nu_e^t|\pi)} \\ \times \sum \begin{bmatrix} S^{0:t-1} \\ {}_{[2:L]}C^t \\ {}_{[2:L]}Z^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha^t \\ {}_{[2:L]}\beta^t \\ \lambda^t \end{bmatrix} \begin{bmatrix} \prod_{j=1}^{t} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \begin{bmatrix} P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{bmatrix} \\ \times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i) \end{bmatrix} \\ \times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi) \end{bmatrix} \end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_e^t|\pi)} \\
\times \sum \begin{bmatrix} S^{0:t-1} \\ {}_{[2:L]}C^t \\ {}_{[2:L]}Z^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha^t \\ {}_{[2:L]}\beta^t \\ \lambda^t \end{bmatrix}
\begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{L} \begin{bmatrix}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{bmatrix} \\
\times P(M^t|\pi) \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\
\prod_{j=1}^{t-1} \begin{bmatrix}
\prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \begin{bmatrix}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{bmatrix} \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{bmatrix} \\
\times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi)
\end{bmatrix}
\right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[1]}\nu_e^t|\pi)} \\
\times \sum_{\left[\ S^{0:t-1}\ \right]} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\sum \left[\begin{array}{l} {}_{[2:L]}C^t \\ {}_{[2:L]}Z^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha^t \\ {}_{[2:L]}\beta^t \\ \lambda^t \end{array} \right] \left[\begin{array}{l}
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right] \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi)
\end{array} \right]
\end{array} \right|
$$

$$
\sum_{\left[\begin{array}{l} {}_{[2:L]}C^t \\ {}_{[2:L]}Z^t \\ {}_{[2:L]}B^t \\ M^t \\ {}_{[2:L]}\alpha^t \\ {}_{[2:L]}\beta^t \\ \lambda^t \end{array} \right]} \left[\begin{array}{l}
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right]
$$

$$
= \sum_{{}_{[2:L]}B^t} \left[\begin{array}{l} \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^t|\pi) \\ \times \prod_{i=1}^{N_i} \sum_{{}_{[k]}\beta_i^t} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \sum_{{}_{[k]}C^t} \left[\begin{array}{l} P({}_{[k]}C^t|\pi) \\ \prod_{i=1}^{N_i} \sum_{{}_{[k]}\alpha_i^t} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} \sum_{{}_{[k]}Z_i^t} P({}_{[k]}Z_i^t|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array} \right] \end{array} \right] \\ \times \sum_{M^t} \left[\begin{array}{l} P(M^t|\pi) \\ \prod_{i=1}^{N_i} \sum_{\lambda_i^t} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{array} \right] \end{array} \right]
$$

$$
\begin{aligned}
& P(S^t|{}_{[1]}\nu_e^t \wedge \pi) \\
= & \left| \begin{array}{l} \frac{1}{P({}_{[1]}\nu_e^t|\pi)} \\ \times P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}z_i^t|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \times \sum_{\left[S^{0:t-1} \right]} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\ \prod_{j=1}^{t-1} \left[\begin{array}{l} \prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{array} \right] \\ \times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i) \end{array} \right] \\ \times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi) \end{array} \right] \end{array} \right. \\
= & \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_e^t|\pi)} \\ \times P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}z_i^t|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \times P(S^t|\nu_p^t \wedge \pi) \end{array} \right.
\end{aligned}
$$

Estimation at level $2 \leq h \leq L$

$$
{}_{[h]}\nu_e^t = \left[\begin{array}{l} {}_{[1:h]}c^{0:t} \wedge {}_{[1:h]}\alpha^{0:t} \wedge {}_{[h+1:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha^{0:t-1} \wedge \\ {}_{[1:h]}b^{0:t} \wedge {}_{[1:h]}\beta^{0:t} \wedge {}_{[h+1:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta^{0:t-1} \wedge \\ {}_{[1:h]}z^{0:t} \wedge {}_{[h+1:L]}z^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \end{array} \right]
$$

$$
\begin{aligned}
&P(S^t|_{[h]}\nu_e^t \wedge \pi) \\
&\quad = P(S^t| \begin{bmatrix} {}_{[1:h]}c^{0:t} \wedge {}_{[1:h]}\alpha^{0:t} \wedge {}_{[h+1:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha^{0:t-1} \wedge \\ {}_{[1:h]}b^{0:t} \wedge {}_{[1:h]}\beta^{0:t} \wedge {}_{[h+1:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta^{0:t-1} \wedge \\ {}_{[1:h]}z^{0:t} \wedge {}_{[h+1:L]}z^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi \end{bmatrix})
\end{aligned}
$$

$$
\begin{aligned}
&P(S^t|_{[h]}\nu_e^t \wedge \pi) \\
&= \left| \begin{array}{l} \frac{1}{P({}_{[h]}\nu_e^t|\pi)} \\ \times \sum \begin{bmatrix} S^{0:t-1} \\ {}_{[h+1:L]}C^t \\ {}_{[h+1:L]}Z^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha^t \\ {}_{[h+1:L]}\beta^t \\ \lambda^t \end{bmatrix} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\ \times P({}_{[1]}B^t|\pi) \times P({}_{[1]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{k=2}^{L} \begin{bmatrix} P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{bmatrix} \\ \times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \\ \prod_{j=1}^{t-1} \begin{bmatrix} \prod_{i=1}^{N_i} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\ \times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\ \prod_{k=2}^{L} \begin{bmatrix} P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i) \end{bmatrix} \\ \times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i) \end{bmatrix} \\ \times P(S^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z^0{}_{[1:L]}\beta^0{}_{[1:L]}\alpha^0\lambda^0|\pi) \end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P({}_{[h]}\nu_e^t|\pi)} \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times \sum \left[\begin{array}{l}
{}_{[h+1:L]}C^t \\
{}_{[h+1:L]}Z^t \\
{}_{[h+1:L]}B^t \\
M^t \\
{}_{[h+1:L]}\alpha^t \\
{}_{[h+1:L]}\beta^t \\
\lambda^t
\end{array} \right] \left[\begin{array}{l}
\prod_{k=h+1}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i)
\end{array} \right] \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0 {}_{[1:L]}B^0 {}_{[1:L]}C^0 {}_{[1:L]}Z^0 {}_{[1:L]}\beta^0 {}_{[1:L]}\alpha^0 \lambda^0|\pi)
\end{array} \right]
\end{array} \right|
$$

Summing over variables in lower levels results in one.

$$\sum_{\begin{bmatrix} {}_{[h+1:L]}C^t \\ {}_{[h+1:L]}Z^t \\ {}_{[h+1:L]}B^t \\ M^t \\ {}_{[h+1:L]}\alpha^t \\ {}_{[h+1:L]}\beta^t \\ \lambda^t \end{bmatrix}} \begin{bmatrix} \prod_{k=h+1}^{L} \begin{bmatrix} P({}_{[k]}B^t|\pi) \times P({}_{[k]}C^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}Z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{bmatrix} \\ \times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{bmatrix}$$

$$= \sum_{{}_{[h+1:L]}B^t} \begin{bmatrix} \prod_{k=h+1}^{L} \begin{bmatrix} P({}_{[k]}B^t|\pi) \times \prod_{i=1}^{N_i} \sum_{{}_{[k]}\beta_i^t} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \sum_{{}_{[k]}C^t} \begin{bmatrix} P({}_{[k]}C^t|\pi) \\ \prod_{i=1}^{N_i} \sum_{{}_{[k]}\alpha_i^t} P({}_{[k]}\alpha_i^t|{}_{[k]}C^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}B^t\pi_i) \\ \prod_{i=1}^{N_i} \sum_{{}_{[k]}Z_i^t} P({}_{[k]}Z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{bmatrix} \end{bmatrix} \\ \times \sum_{M^t} \begin{bmatrix} P(M^t|\pi) \\ \times \prod_{i=1}^{N_i} \sum_{\lambda_i^t} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}B^t \wedge S_i^t\pi_i) \end{bmatrix} \end{bmatrix}$$

Estimation question can the be written in function of prediction:

$$P(S^t|_{[t]}\nu_e^{\wedge}\pi)$$

$$
\begin{aligned}
&= \left|\begin{array}{l}
\frac{1}{P({}_{[h]}\nu_e^t|\pi)} \\
\times P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array}\right] \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array}\right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array}\right] \\
\times P(S^0 M^0{}_{[1:L]} B^0{}_{[1:L]} C^0{}_{[1:L]} Z^0{}_{[1:L]} \beta^0{}_{[1:L]} \alpha^0 \lambda^0|\pi)
\end{array}\right]
\end{array}\right| \\
&= \left|\begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_e^t|\pi)} \\
\times P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array}\right] \\
\times P(S^t|\nu_p^t \wedge \pi)
\end{array}\right|
\end{aligned}
\tag{E.12}
$$

Particularly for level $h = L$, estimation is:

$$P(S^t|_{[L]}\nu_e^t \wedge \pi)$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_e^t|\pi)} \\
\times P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P(S^t|\nu_p^t \wedge \pi)
\end{array} \right| \tag{E.13}
$$

And ${}_{[L]}\nu_e^t$ in this case is:

$$
{}_{[L]}\nu_e^t = \left[\begin{array}{l}
{}_{[1:L]}c^{0:t} \wedge {}_{[1:L]}\alpha^{0:t} \wedge {}_{[1:L]}z^{0:t} \wedge \\
{}_{[1:L]}b^{0:t} \wedge {}_{[1:L]}\beta^{0:t} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi
\end{array} \right]
$$

E.2.6 Motor Commands

$$
\nu_m^t = \; {}_{[1:L]}c^{0:t} \wedge {}_{[1:L]}\alpha^{0:t} \wedge {}_{[1:L]}z^{0:t} \wedge {}_{[1:L]}b^{0:t} \wedge {}_{[1:L]}\beta^{0:t} \wedge m^{0:t-1} \wedge \lambda^{0:t}
$$

$$
\begin{array}{l}
P(M^t|\nu_m^t \wedge \pi) \\
\quad = P(M^t|\; {}_{[1:L]}c^{0:t} \wedge {}_{[1:L]}\alpha^{0:t} \wedge {}_{[1:L]}z^{0:t} \wedge {}_{[1:L]}b^{0:t} \wedge {}_{[1:L]}\beta^{0:t} \wedge m^{0:t-1} \wedge \lambda^{0:t} \wedge \pi \;)
\end{array}
$$

Motor question can then be written in function of estimation at level L detailled in equation E.13.

$$
\begin{array}{l}
P(M^t|\nu_m^t \wedge \pi) \\
= \left| \begin{array}{l}
\frac{1}{P(\nu_m^t|\pi)} \\
\times \sum_{S^{0:t}} \left[\prod_{j=1}^{t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \right] \\
\times P(S^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z^0{}_{[1:L]}\beta^0{}_{[1:L]}\alpha^0\lambda^0|\pi) \Bigg]
\end{array} \right|
\end{array}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_m^t|\pi)} \\
\times \sum_{S^{0:t}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi) \times \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}c^t|\pi) \times \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times \prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\times \prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi) \times \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}c^t|\pi) \times \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times \prod_{i=1}^{N_i} P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}b^t \wedge S_i^t\pi_i) \\
\times \prod_{j=1}^{t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,|S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,|S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,|S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0{}_{[1:L]}B^0{}_{[1:L]}C^0{}_{[1:L]}Z^0{}_{[1:L]}\beta^0{}_{[1:L]}\alpha^0\lambda^0|\pi)
\end{array} \right] \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_m^t|\pi)} \\
\times \sum_{S^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi) \times \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}c^t|\pi) \times \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times \prod_{i=1}^{N_i} P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\times \prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi) \times \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}c^t|\pi) \times \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times \prod_{i=1}^{N_i} P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}b^t \wedge S_i^t\pi_i) \\
\times \sum_{S^{0:t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,\big|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times \prod_{j=1}^{t-1} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^j \,\big|\, S_i^{j-1} \wedge M^{j-1} \wedge \pi_i) \\
\times P({}_{[1]}B^j|\pi) \times P({}_{[1]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^j|{}_{[1]}B^j \wedge S_i^j \wedge {}_{[1]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^j|{}_{[1]}C^j \wedge S_i^j \wedge {}_{[1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}Z_i^j \,\big|\, S_i^j \wedge {}_{[1]}C^j \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^j|\pi) \times P({}_{[k]}C^j|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^j|{}_{[k]}B^j \wedge S_i^j \wedge {}_{[k-1]}B^j \wedge {}_{[k]}B^{j-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^j|{}_{[k]}C^j \wedge S_i^j \wedge {}_{[k]}B^j \wedge {}_{[k-1]}B^j\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^j \,\big|\, S_i^j \wedge {}_{[k]}C^j \wedge \pi_i)
\end{array} \right] \\
\times P(M^j|\pi) \prod_{i=1}^{N_i} P(\lambda_i^j|M^j \wedge M^{j-1} \wedge {}_{[L]}B^j \wedge S_i^j\pi_i)
\end{array} \right] \\
\times P(S^0 M^0{}_{[1:L]} B^0{}_{[1:L]} C^0{}_{[1:L]} Z^0{}_{[1:L]} \beta^0{}_{[1:L]} \alpha^0 \lambda^0|\pi)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P({}_{[L]}\nu_e^t|\pi)}{P(\nu_m^t|\pi)} \\
\times \sum_{S^t} \left[\begin{array}{l}
P(M^t|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}b^t \wedge S_i^t\pi_i) \\
\times P(S^t|\nu_m^t \wedge \pi)
\end{array} \right]
\end{array} \right|
$$

E.2.7 Prediction Loop

Taking final form of prediction, we can write it in relation to estimation at level L at previous time step.

$$P(S^t|\nu_p^t \wedge \pi)$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^{t-1}|\pi)}{P(\nu_p^{t}|\pi)} \\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \left| S_i^{t-1} \wedge m^{t-1} \wedge \pi_i \right.) \\
\times P({}_{[1]}b^{t-1}|\pi) \times P({}_{[1]}c^{t-1}|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^{t-1}|{}_{[1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^{t-1}|{}_{[1]}C^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^{t-1} \left| S_i^{t-1} \wedge {}_{[1]}c^{t-1} \wedge \pi_i \right.) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^{t-1}|\pi) \times P({}_{[k]}c^{t-1}|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^{t-1}|{}_{[k]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[k-1]}b^{t-1} \wedge {}_{[k]}b^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^{t-1}|{}_{[k]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^{t-1} \left| S_i^{t-1} \wedge {}_{[k]}c^{t-1} \wedge \pi_i \right.)
\end{array} \right] \\
\times P(m^{t-1}|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^{t-1}|m^{t-1} \wedge m^{t-2} \wedge {}_{[L]}b^{t-1} \wedge S_i^{t-1}\pi_i) \\
\times P(S^{t-1}|\nu_p^{t-1} \wedge \pi)
\end{array} \right]
\end{array} \right.
$$

$$
= \left| \begin{array}{l}
\frac{P({}_{[L]}\nu_e^{t-1}|\pi)}{P(\nu_p^{t}|\pi)} \\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \left| S_i^{t-1} \wedge m^{t-1} \wedge \pi_i \right.) \\
\times P(m^{t-1}|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^{t-1}|m^{t-1} \wedge m^{t-2} \wedge {}_{[L]}b^{t-1} \wedge S_i^{t-1}\pi_i) \\
\times P(S^{t-1}|_{[L]}\nu_e^{t-1} \wedge \pi)
\end{array} \right]
\end{array} \right.
$$

E.2.8 Attention and Behaviour at level h, with $2 \leq h \leq L$

Lets look at expression for attention expression at a general level h, with $2 \leq h \leq L$:

$$
P({}_{[h]}C^t \wedge {}_{[h]}B^t|_{[h]}\nu_a^t \wedge \pi)
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^{t}|\pi)}{P({}_{[h]}\nu_a^{t}|\pi)} \\
\times \sum_{S^{t}} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \left| S_i^t \wedge {}_{[1]}c^t \wedge \pi_i \right.) \\
\prod_{k=2}^{h-1} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi) \times P({}_{[k]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t \left| S_i^t \wedge {}_{[k]}c^t \wedge \pi_i \right.)
\end{array} \right] \\
\times P({}_{[h]}B^t|\pi) \times P({}_{[h]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[h]}\beta_i^t|{}_{[h]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[h]}\alpha_i^t|{}_{[h]}C^t \wedge S_i^t \wedge {}_{[h]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P(S^t|\nu_p^t \wedge \pi)
\end{array} \right]
\end{array} \right.
$$

It is possible to see that it function of prediction, but it is possible to go farther and write it in relation to estimation at previous level $h-1$. Estimation is as stated in equation E.12, and we find:

$$
\begin{aligned}
&P({}_{[h]}C^t \wedge {}_{[h]}B^t|_{[h]}\nu_a^t \wedge \pi) \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_a^t|\pi)} \times \frac{P({}_{[h]}\nu_e^t|\pi)}{P(\nu_p^t|\pi)} \\ \times \sum_{S^t} \left[\begin{array}{l} P({}_{[h]}B^t|\pi) \times P({}_{[h]}C^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[h]}\beta_i^t|_{[h]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[h]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[h]}\alpha_i^t|_{[h]}C^t \wedge S_i^t \wedge {}_{[h]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\ \times P(S^t|E_{pres}^{h-1}t \wedge \pi) \end{array} \right] \end{array} \right.
\end{aligned}
$$

The same principle applies for behaviour expression at a general level h, with $2 \leq h \leq L$:

$$
\begin{aligned}
&P({}_{[h]}B^t{}_{[h]}\nu_b^t \wedge \pi) \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_b^t|\pi)} \\ \times \sum_{S^t} \left[\begin{array}{l} P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{k=2}^{h} \left[\begin{array}{l} P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}B^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i) \end{array} \right] \\ \times P(S^t|\nu_p^t \wedge \pi) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_b^t|\pi)} \\ \times \sum_{S^t} \left[\begin{array}{l} P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{k=2}^{h-1} \left[\begin{array}{l} P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}B^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i) \end{array} \right] \\ \times P({}_{[h]}B^t|\pi) \times P({}_{[h]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[h]}\beta_i^t|_{[h]}B^t \wedge S_i^t \wedge {}_{[h-1]}B^t \wedge {}_{[h]}B^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[h]}\alpha_i^t|_{[h]}c^t \wedge S_i^t \wedge {}_{[h]}B^t \wedge {}_{[h-1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[h]}z_i^t\,|S_i^t \wedge {}_{[h]}c^t \wedge \pi_i)P(S^t|\nu_p^t \wedge \pi) \end{array} \right] \end{array} \right.
\end{aligned}
$$

$$= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_b^t|\pi)} \times \frac{P({}_{[h]}\nu_e^t|\pi)}{P(\nu_p^t|\pi)} \\ \times \sum_{S^t} \left[\begin{array}{l} P({}_{[h]}B^t|\pi) \times P({}_{[h]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[h]}\beta_i^t|_{[h]}B^t \wedge S_i^t \wedge {}_{[h-1]}b^t \wedge {}_{[h]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[h]}\alpha_i^t|_{[h]}c^t \wedge S_i^t \wedge {}_{[h]}B^t \wedge {}_{[h-1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[h]}z_i^t|S_i^t \wedge {}_{[h]}c^t \wedge \pi_i)P(S^t|_{[h-1]}\nu_e^t \wedge \pi) \end{array} \right] \end{array} \right|$$

Last, it is also possible to write estimation at level h in relation to estimation at level $h-1$.

$$\begin{array}{l} P(S^t|_{[t]}\nu_e^{\wedge}\pi) \\ = \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_e^t|\pi)} \\ \times P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}z_i^t|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{k=2}^{h} \left[\begin{array}{l} P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}z_i^t|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i) \end{array} \right] \\ \times P(S^t|\nu_p^t \wedge \pi) \end{array} \right| \\ = \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_e^t|\pi)} \\ \times P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}z_i^t|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{k=2}^{h-1} \left[\begin{array}{l} P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}z_i^t|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i) \end{array} \right] \\ \times P({}_{[h]}b^t|\pi) \times P({}_{[h]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[h]}\beta_i^t|_{[h]}b^t \wedge S_i^t \wedge {}_{[h-1]}b^t \wedge {}_{[h]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[h]}\alpha_i^t|_{[h]}c^t \wedge S_i^t \wedge {}_{[h]}b^t \wedge {}_{[h-1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[h]}z_i^t|S_i^t \wedge {}_{[h]}c^t \wedge \pi_i)P(S^t|\nu_p^t \wedge \pi) \end{array} \right| \\ = \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_e^t|\pi)} \times \frac{P({}_{[h]}\nu_e^t|\pi)}{P(\nu_p^t|\pi)} \\ \times P({}_{[h]}b^t|\pi) \times P({}_{[h]}c^t|\pi) \\ \prod_{i=1}^{N_i} P({}_{[h]}\beta_i^t|_{[h]}b^t \wedge S_i^t \wedge {}_{[h-1]}b^t \wedge {}_{[h]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[h]}\alpha_i^t|_{[h]}c^t \wedge S_i^t \wedge {}_{[h]}b^t \wedge {}_{[h-1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[h]}z_i^t|S_i^t \wedge {}_{[h]}c^t \wedge \pi_i)P(S^t|E_{pres}^{h-1}t \wedge \pi) \end{array} \right| \end{array}$$

E.3 Relations Between Elementary and Global Filter

Prediction

We aim to show that prediction in global filter is proportional to product of prediction in elementary filters. We begin assuming that this is true at time $t-1$, as in:

$$
\begin{aligned}
& P(S^{t-1}|\nu_p^t \wedge \pi) = \\
& \quad = \quad \kappa_{t-2} \times \textstyle\prod_{i=1}^{N_i} P(S_i^{t-1}|\nu_p^t \wedge \pi_i)
\end{aligned}
$$

Developing prediction at time t, we find:

$$
\begin{aligned}
& P(S^t|\nu_p^t \wedge \pi) \\
= & \left| \begin{array}{l}
\frac{P(\nu_p^{t-1}|\pi)}{P(\nu_p^t|\pi)} \\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^{t-1}|\pi) \times P({}_{[1]}c^{t-1}|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^{t-1}|{}_{[1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^{t-1} \,|S_i^{t-1} \wedge {}_{[1]}c^{t-1} \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^{t-1}|\pi) \times P({}_{[k]}c^{t-1}|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^{t-1}|{}_{[k]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[k-1]}b^{t-1} \wedge {}_{[k]}b^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^{t-1}|{}_{[k]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^{t-1} \,|S_i^{t-1} \wedge {}_{[k]}c^{t-1} \wedge \pi_i)
\end{array} \right] \\
\times P(m^{t-1}|\pi) \times \prod_{i=1}^{N_i} P(\lambda_i^{t-1}|m^{t-1} \wedge m^{t-2} \wedge {}_{[L]}b^{t-1} \wedge S_i^{t-1}\pi_i) \\
\times P(S^{t-1}|\nu_p^t \wedge \pi)
\end{array} \right]
\end{array} \right. \\
= & \left| \begin{array}{l}
\frac{P(\nu_p^{t-1}|\pi)}{P(\nu_p^t|\pi)} \times P(m^{t-1}|\pi) \times \prod_{k=1}^{L} \left[P({}_{[k]}b^{t-1}|\pi) \times P({}_{[k]}c^{t-1}|\pi) \right] \\
\times \sum_{S^{t-1}} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^{t-1}|{}_{[1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^{t-1}|{}_{[1]}C^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^{t-1} \,|S_i^{t-1} \wedge {}_{[1]}c^{t-1} \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^{t-1}|{}_{[k]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[k-1]}b^{t-1} \wedge {}_{[k]}b^{t-2}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^{t-1}|{}_{[k]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^{t-1} \,|S_i^{t-1} \wedge {}_{[k]}c^{t-1} \wedge \pi_i)
\end{array} \right] \\
\prod_{i=1}^{N_i} P(\lambda_i^{t-1}|m^{t-1} \wedge m^{t-2} \wedge {}_{[L]}b^{t-1} \wedge S_i^{t-1}\pi_i) \\
\times \kappa_{t-2} \times \prod_{i=1}^{N_i} P(S_i^{t-1}|\nu_p^t \wedge \pi_i)
\end{array} \right]
\end{array} \right.
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^{t-1}|\pi)}{P(\nu_p^{t}|\pi)} \times \kappa_{t-2} \times \frac{P(m^{t-1}|\pi) \prod_{k=1}^{L} \left[P({}_{[k]}b^{t-1}|\pi) \times P({}_{[k]}c^{t-1}|\pi) \right]}{\prod_{i=1}^{N_i} \left[P(m^{t-1}|\pi_i) \prod_{k=1}^{L} \left[P({}_{[k]}b^{t-1}|\pi_i) \times P({}_{[k]}c^{t-1}|\pi_i) \right] \right]} \\
\times \sum_{S^{t-1}} \left[\prod_{i=1}^{N_i} \left[\begin{array}{l}
P(S_i^t \big| S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^{t-1}|\pi_i) \times P({}_{[1]}c^{t-1}|\pi_i) \\
P({}_{[1]}\beta_i^{t-1}|_{[1]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-2}\pi_i) \\
\times P({}_{[1]}\alpha_i^{t-1}|_{[1]}C^{t-1} \wedge S_i^{t-1} \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}z_i^{t-1} \big| S_i^{t-1} \wedge {}_{[1]}c^{t-1} \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^{t-1}|\pi_i) \times P({}_{[k]}c^{t-1}|\pi_i) \\
P({}_{[k]}\beta_i^{t-1}|_{[k]}b^{t-1} \wedge S_i^{t-1} \wedge {}_{[k-1]}b^{t-1} \wedge {}_{[k]}b^{t-2}\pi_i) \\
\times P({}_{[k]}\alpha_i^{t-1}|_{[k]}c^{t-1} \wedge S_i^{t-1} \wedge {}_{[k]}b^{t-1} \wedge {}_{[k-1]}b^{t-1}\pi_i) \\
\times P({}_{[k]}z_i^{t-1} \big| S_i^{t-1} \wedge {}_{[k]}c^{t-1} \wedge \pi_i)
\end{array} \right] \\
\times P(m^{t-1}|\pi_i) \\
\times P(\lambda_i^{t-1}|m^{t-1} \wedge m^{t-2} \wedge {}_{[L]}b^{t-1} \wedge S_i^{t-1}\pi_i) \\
\times P(S_i^{t-1}|\nu_p^t \wedge \pi_i)
\end{array} \right] \right] \right.
\end{array} \right.
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^{t-1}|\pi)}{P(\nu_p^{t}|\pi)} \times \kappa_{t-2} \times \frac{P(m^{t-1}|\pi) \prod_{k=1}^{L} \left[P({}_{[k]}b^{t-1}|\pi) \times P({}_{[k]}c^{t-1}|\pi) \right]}{\prod_{i=1}^{N_i} \left[P(m^{t-1}|\pi_i) \prod_{k=1}^{L} \left[P({}_{[k]}b^{t-1}|\pi_i) \times P({}_{[k]}c^{t-1}|\pi_i) \right] \right]} \\
\prod_{i=1}^{N_i} \left[\; P(S_i^t|\nu_p^t \wedge \pi_i) \; \right]
\end{array} \right. \tag{E.14}
$$

from the expression above, we find how to calculate κ_{t-1}:

$$
\begin{aligned}
\kappa_{t-1} &= \\
&= \frac{P(\nu_p^{t-1}|\pi)}{P(\nu_p^{t}|\pi)} \times \kappa_{t-2} \times \frac{P(m^{t-1}|\pi) \prod_{k=1}^{L} \left[P({}_{[k]}b^{t-1}|\pi) \times P({}_{[k]}c^{t-1}|\pi) \right]}{\prod_{i=1}^{N_i} \left[P(m^{t-1}|\pi_i) \prod_{k=1}^{L} \left[P({}_{[k]}b^{t-1}|\pi_i) \times P({}_{[k]}c^{t-1}|\pi_i) \right] \right]}
\end{aligned}
$$

We have then shown that if prediction at global filter is proportional to product of prediction at elementary filters at time $t-1$, so this is also true at time t. It is sufficient to show that the proportionality exists at time $t=1$, and proof is complete.

Right part of prediction question at $t=1$ for an elementary filter is:

$$
\nu_p^1 = {}_{[1:L]}z_i^0 \wedge {}_{[1:L]}c^0 \wedge {}_{[1:L]}b^0 \wedge m^0 \wedge \lambda_i^0 \wedge {}_{[1:L]}\beta_i^0 \wedge {}_{[1:L]}\alpha_i^0
$$

$$
\begin{aligned}
&P(S_i^1|\nu_p^1 \wedge \pi_i) \\
&\quad P(S_i^1|_{[1:L]}z_i^0 \wedge {}_{[1:L]}c^0 \wedge {}_{[1:L]}b^0 \wedge m^0 \wedge \lambda_i^0 \wedge {}_{[1:L]}\beta_i^0 \wedge {}_{[1:L]}\alpha_i^0 \wedge \pi_i)
\end{aligned}
$$

And prediction expression for an elementary filter at time $t=1$ is shown below.

$$
\begin{aligned}
P(S_i^1|\nu_p^1 \wedge \pi_i) &= \\
&= \left| \begin{array}{l}
\frac{1}{P(\nu_p^1|\pi_i)} \\
\sum \begin{bmatrix} S_i^0 \\ {}_{[1:L]}Z_i^1 \\ {}_{[1:L]}C^1 \\ {}_{[1:L]}B^1 \\ M^1 \\ \lambda_i^1 \\ {}_{[1:L]}\beta_i^1 \\ {}_{[1:L]}\alpha_i^1 \end{bmatrix}
\begin{bmatrix}
P(S_i^0\,|\pi_i) \times P(m^0\,|\pi_i) \\
\prod_{k=1}^{L} \begin{bmatrix} P({}_{[k]}b^0\,|\pi_i) \times P({}_{[k]}c^0|\pi_i) \\ \times P({}_{[k]}\beta_i^0|{}_{[k]}b^0 \wedge S_i^0\pi_i)\times \\ P({}_{[k]}\alpha_i^0|{}_{[k]}c^0 \wedge S_i^0 \wedge {}_{[k]}b^0\pi_i)\times \\ P({}_{[k]}z_i^0\,|S_i^0 \wedge {}_{[k]}c^0 \wedge \pi_i) \end{bmatrix} \\
\times P(S_i^1\,|S_i^0 \wedge m^0 \wedge \pi_i)\times \\
P({}_{[1]}B^1|\pi_i) \times P({}_{[1]}C^1|\pi_i) \\
\times P({}_{[1]}\beta_i^1|{}_{[1]}B^1 \wedge S_i^1 \wedge {}_{[1]}B^0\pi_i) \\
\times P({}_{[1]}\alpha_i^1|{}_{[1]}C^1 \wedge S_i^1 \wedge {}_{[1]}B^1\pi_i) \\
\times P({}_{[1]}Z_i^1\,|S_i^1 \wedge {}_{[1]}C^1 \wedge \pi_i) \\
\prod_{k=2}^{L} \begin{bmatrix} P({}_{[k]}B^1|\pi_i) \times P({}_{[k]}C^1|\pi_i) \\ \times P({}_{[k]}\beta_i^1|{}_{[k]}B^1 \wedge {}_{[k-1]}B^1 \wedge S_i^1 \wedge {}_{[k]}b^0\pi_i) \\ \times P({}_{[k]}\alpha_i^1|{}_{[k]}C^1 \wedge S_i^1 \wedge {}_{[k]}B^1 \wedge {}_{[k-1]}B^1\pi_i) \\ \times P({}_{[k]}Z_i^1\,|S_i^1 \wedge {}_{[k]}C^1 \wedge \pi_i) \end{bmatrix} \\
\times P(M^1|\pi_i) \times P(\lambda_i^1|M^1 \wedge m^0 \wedge {}_{[L]}B^j \wedge S_i^1\pi_i)
\end{array} \right] \\
&= \left| \begin{array}{l}
\sum_{S_i^0} \left[\begin{array}{l}
\frac{1}{P(\nu_p^1|\pi_i)} \\
P(S_i^0\,|\pi_i) \times P(m^0\,|\pi_i) \\
\prod_{k=1}^{L} \begin{bmatrix} P({}_{[k]}b^0\,|\pi_i) \times P({}_{[k]}c^0|\pi_i) \\ \times P({}_{[k]}\beta_i^0|{}_{[k]}b^0 \wedge S_i^0\pi_i) \\ \times P({}_{[k]}\alpha_i^0|{}_{[k]}c^0 \wedge S_i^0 \wedge {}_{[k]}b^0\pi_i) \\ \times P({}_{[k]}z_i^0\,|S_i^0 \wedge {}_{[k]}c^0 \wedge \pi_i) \end{bmatrix} \\
\times P(S_i^1\,|S_i^0 \wedge m^0 \wedge \pi_i) \\
\times \sum \begin{bmatrix} {}_{[1:L]}Z_i^1 \\ {}_{[1:L]}C^1 \\ {}_{[1:L]}B^1 \\ M^1 \\ \lambda_i^1 \\ {}_{[1:L]}\beta_i^1 \\ {}_{[1:L]}\alpha_i^1 \end{bmatrix}
\begin{bmatrix}
P({}_{[1]}B^1|\pi_i) \times P({}_{[1]}C^1|\pi_i) \\
\times P({}_{[1]}\beta_i^1|{}_{[1]}B^1 \wedge S_i^1 \wedge {}_{[1]}b^0\pi_i) \\
\times P({}_{[1]}\alpha_i^1|{}_{[1]}C^1 \wedge S_i^1 \wedge {}_{[1]}B^1\pi_i) \\
\times P({}_{[1]}Z_i^1\,|S_i^1 \wedge {}_{[1]}C^1 \wedge \pi_i) \\
\prod_{k=2}^{L} \begin{bmatrix} P({}_{[k]}B^1|\pi_i) \times P({}_{[k]}C^1|\pi_i) \\ \times P({}_{[k]}\beta_i^1|{}_{[k]}B^1 \wedge {}_{[k-1]}B^1 \wedge S_i^1 \wedge {}_{[k]}b^0\pi_i) \\ \times P({}_{[k]}\alpha_i^1|{}_{[k]}C^1 \wedge S_i^1 \wedge {}_{[k]}B^1 \wedge {}_{[k-1]}B^1\pi_i) \\ \times P({}_{[k]}Z_i^1\,|S_i^1 \wedge {}_{[k]}C^1 \wedge \pi_i) \end{bmatrix} \\
\times P(M^1|\pi_i) \times P(\lambda_i^1|M^1 \wedge m^0 \wedge {}_{[L]}B^j \wedge S_i^1\pi_i)
\end{array} \right]
\end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^1|\pi_i)} \times P(m^0\,|\pi_i) \\
\times \sum_{S_i^0} \left[\begin{array}{l}
P(S_i^0\,|\pi_i) \prod_{k=1}^{L} \left[\begin{array}{l}
P({}_{[k]}b^0\,|\pi_i) \times P({}_{[k]}c^0|\pi_i) \\
\times P({}_{[k]}\beta_i^0|{}_{[k]}b^0 \wedge S_i^0\pi_i) \\
\times P({}_{[k]}\alpha_i^0|{}_{[k]}c^0 \wedge S_i^0 \wedge {}_{[k]}b^0\pi_i) \\
\times P({}_{[k]}z_i^0\,|S_i^0 \wedge {}_{[k]}c^0 \wedge \pi_i)
\end{array} \right] \\
\times P(S_i^1\,|S_i^0 \wedge m^0 \wedge \pi_i)
\end{array} \right]
\end{array} \right|
$$

Now, we want to show the relation between above expression and Global filter prediction at time $t = 1$. Right part of Global filter prediction question is then:

$$
\nu_p^1 = {}_{[1:L]}z^0 \wedge {}_{[1:L]}c^0 \wedge {}_{[1:L]}b^0 \wedge m^0 \wedge \lambda^0 \wedge {}_{[1:L]}\beta^0 \wedge {}_{[1:L]}\alpha^0
$$

$$
\begin{array}{l}
P(S^1|\nu_p^1 \wedge \pi) \\
\quad P(S^1|_{[1:L]}z^0 \wedge {}_{[1:L]}c^0 \wedge {}_{[1:L]}b^0 \wedge m^0 \wedge \lambda^0 \wedge {}_{[1:L]}\beta^0 \wedge {}_{[1:L]}\alpha^0 \wedge \pi)
\end{array}
$$

And prediction expression at Global filter for time $t = 1$ is:

$$
P(S^1|\nu_p^1 \wedge \pi) =
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^1|\pi)} \\
\times \sum \left[\begin{array}{l}
S^0 \\
{}_{[1:L]}Z^1 \\
{}_{[1:L]}C^1 \\
{}_{[1:L]}B^1 \\
M^1 \\
\lambda^1 \\
{}_{[1:L]}\beta^1 \\
{}_{[1:L]}\alpha^1
\end{array} \right]
\left[\begin{array}{l}
P(S^0\,|\pi) \times P(m^0\,|\pi) \\
\prod_{k=1}^{L} \left[\begin{array}{l}
P({}_{[k]}b^0\,|\pi) \times P({}_{[k]}c^0|\pi) \\
\times P({}_{[k]}\beta^0|S^0 \wedge {}_{[k]}b^0\pi) \times \\
P({}_{[k]}\alpha^0|_{[k]}c^0 \wedge S^0 \wedge {}_{[k]}b^0\pi) \times \\
P({}_{[k]}z^0\,|S^0 \wedge {}_{[k]}c^0\pi)
\end{array} \right] \\
\prod_{i=1}^{N_i} P(S_i^1\,|S_i^0 \wedge m^0 \wedge \pi_i) \\
\times P({}_{[1]}B^1|\pi) \times P({}_{[1]}C^1|\pi) \\
\times \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^1|_{[1]}B^1 \wedge S_i^1 \wedge {}_{[1]}b^0\pi_i) \\
\times \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^1|_{[1]}C^1 \wedge S_i^1 \wedge {}_{[1]}B^1\pi_i) \\
\times \prod_{i=1}^{N_i} P({}_{[1]}Z_i^1\,|S_i^1 \wedge {}_{[1]}C^1 \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}B^1|\pi) \times P({}_{[k]}C^1|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^1|_{[k]}B^1 \wedge S_i^1 \wedge {}_{[k-1]}B^1 \wedge {}_{[k]}b^0\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^1|_{[k]}C^1 \wedge S_i^1 \wedge {}_{[k]}B^1 \wedge {}_{[k-1]}B^1\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^1\,|S_i^1 \wedge {}_{[k]}C^1 \wedge \pi_i)
\end{array} \right] \\
\times P(M^1|\pi) \prod_{i=1}^{N_i} P(\lambda_i^1|M^1 \wedge m^0 \wedge {}_{[L]}B^1 \wedge S_i^1\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^1|\pi)} \\
\times \sum S^0 \left[\begin{array}{l}
P(S^0\,|\pi) \times P(m^0\,|\pi) \\
\prod_{k=1}^{L} \left[\begin{array}{l} P({}_{[k]}b^0\,|\pi) \times P({}_{[k]}c^0|\pi) \\ \times P({}_{[k]}\beta^0|S^0 \wedge {}_{[k]}b^0\pi) \times \\ P({}_{[k]}\alpha^0|{}_{[k]}c^0 \wedge S^0 \wedge {}_{[k]}b^0\pi) \times \\ P({}_{[k]}z^0\,|S^0 \wedge {}_{[k]}c^0\pi) \end{array} \right] \\
\prod_{i=1}^{N_i} P(S_i^1\,|S_i^0 \wedge m^0 \wedge \pi_i) \\
\times \sum \left[\begin{array}{l} {}_{[1:L]}Z^1 \\ {}_{[1:L]}C^1 \\ {}_{[1:L]}B^1 \\ M^1 \\ \lambda^1 \\ {}_{[1:L]}\beta^1 \\ {}_{[1:L]}\alpha^1 \end{array} \right] \left[\begin{array}{l}
P({}_{[1]}B^1|\pi) \times P({}_{[1]}C^1|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^1|{}_{[1]}B^1 \wedge S_i^1 \wedge {}_{[1]}b^0\pi_i) \\
\times \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^1|{}_{[1]}C^1 \wedge S_i^1 \wedge {}_{[1]}B^1\pi_i) \\
\times \prod_{i=1}^{N_i} P({}_{[1]}Z_i^1\,|S_i^1 \wedge {}_{[1]}C^1 \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l} P({}_{[k]}B^1|\pi) \times P({}_{[k]}C^1|\pi) \\ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^1|{}_{[k]}B^1 \wedge S_i^1 \wedge {}_{[k-1]}B^1 \wedge {}_{[k]}b^0\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^1|{}_{[k]}C^1 \wedge S_i^1 \wedge {}_{[k]}B^1 \wedge {}_{[k-1]}B^1\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}Z_i^1\,|S_i^1 \wedge {}_{[k]}C^1 \wedge \pi_i) \end{array} \right] \\
\times P(M^1|\pi) \prod_{i=1}^{N_i} P(\lambda_i^1|M^1 \wedge m^0 \wedge {}_{[L]}B^1 \wedge S_i^1\pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^1|\pi)} \times \frac{P(m^0|\pi) \prod_{k=1}^{L} \left[P({}_{[k]}b^0|\pi) \times P({}_{[k]}c^0|\pi) \right]}{\prod_{i=1}^{N_i} \left[P(m^0|\pi_i) \prod_{k=1}^{L} \left[P({}_{[k]}b^0|\pi_i) \times P({}_{[k]}c^0|\pi_i) \right] \right]} \\
\times \sum S^0 \left[\prod_{i=1}^{N_i} \left[\begin{array}{l}
P(S_i^0\,|\pi_i) \times P(m^0\,|\pi_i) \\
\prod_{k=1}^{L} \left[\begin{array}{l} P({}_{[k]}b^0\,|\pi_i) \times P({}_{[k]}c^0|\pi_i) \\ \times P({}_{[k]}\beta_i^0|S_i^0 \wedge {}_{[k]}b^0\pi_i) \\ \times P({}_{[k]}\alpha_i^0|{}_{[k]}c^0 \wedge S_i^0 \wedge {}_{[k]}b^0\pi_i) \\ \times P({}_{[k]}z_i^0\,|S_i^0 \wedge {}_{[k]}c^0\pi_i) \end{array} \right] \\
\times P(S_i^1\,|S_i^0 \wedge m^0 \wedge \pi_i)
\end{array} \right] \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{1}{P(\nu_p^1|\pi)} \times \frac{P(m^0|\pi) \prod_{k=1}^{L} \left[P({}_{[k]}b^0|\pi) \times P({}_{[k]}c^0|\pi) \right]}{\prod_{i=1}^{N_i} \left[P(m^0|\pi_i) \prod_{k=1}^{L} \left[P({}_{[k]}b^0|\pi_i) \times P({}_{[k]}c^0|\pi_i) \right] \right]} \\
\prod_{i=1}^{N_i} \left[\sum S_i^0 \left[\begin{array}{l}
P(S_i^0\,|\pi_i) \times P(m^0\,|\pi_i) \\
\prod_{k=1}^{L} \left[\begin{array}{l} P({}_{[k]}b^0\,|\pi_i) \times P({}_{[k]}c^0|\pi_i) \\ P({}_{[k]}\beta_i^0|S_i^0 \wedge {}_{[k]}b^0\pi_i) \\ \times P({}_{[k]}\alpha_i^0|{}_{[k]}c^0 \wedge S_i^0 \wedge {}_{[k]}b^0\pi_i) \\ \times P({}_{[k]}z_i^0\,|S_i^0 \wedge {}_{[k]}c^0\pi_i) \end{array} \right] \\
\times P(S_i^1\,|S_i^0 \wedge m^0 \wedge \pi_i)
\end{array} \right] \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{\prod_{i=1}^{N_i} [P(\nu_p^1|\pi_i)]}{P(\nu_p^1|\pi)} \times \frac{P(m^0|\pi) \prod_{k=1}^{L} \left[P({}_{[k]}b^0|\pi) \times P({}_{[k]}c^0|\pi) \right]}{\prod_{i=1}^{N_i} \left[P(m^0|\pi_i) \prod_{k=1}^{L} \left[P({}_{[k]}b^0|\pi_i) \times P({}_{[k]}c^0|\pi_i) \right] \right]} \\
\prod_{i=1}^{N_i} \left[\; P(S_i^1|\nu_p^1 \wedge \pi_i) \; \right]
\end{array} \right|
$$

And so, we can find expression for κ_0:

$$
\begin{aligned}
&\kappa_0 \\
&\quad = \frac{\prod_{i=1}^{N_i} P(\nu_p^1|\pi_i)}{P(\nu_p^1|\pi)} \times \frac{P(m^0|\pi) \prod_{k=1}^{L} \left[P({}_{[k]}b^0|\pi) \times P({}_{[k]}c^0|\pi) \right]}{\prod_{i=1}^{N_i} \left[P(m^0|\pi_i) \prod_{k=1}^{L} \left[P({}_{[k]}b^0|\pi_i) \times P({}_{[k]}c^0|\pi_i) \right] \right]}
\end{aligned}
$$

Attention Selection

It is desired to show attention question at level 1 to Global distribution is proportional to the product of attention question at level 1 in the elementary filters.

Right part of attention Selection at level 1 for Global filter is:

$$_{[1]}\nu_a^t = {}_{[1:L]}c^{0:t-1} \wedge {}_{[1:L]}\alpha^{0:t} \wedge {}_{[1:L]}z^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1:L]}\beta^{0:t} \wedge m^{0:t-1} \wedge \lambda^{0:t-1}$$

$$P({}_{[1]}C^t \wedge {}_{[1]}B^t|_{[1]}\nu_a^t \wedge \pi)$$
$$P({}_{[1]}C^t \wedge {}_{[1]}B^t|_{[1:L]}c^{0:t-1} \wedge {}_{[1:L]}\alpha^{0:t} \wedge {}_{[1:L]}z^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1:L]}\beta^{0:t} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi)$$

Developing question for attention (with behaviour associated) at level 1 for global filter, we find:

$$\begin{aligned}
&P({}_{[1]}C^t \wedge {}_{[1]}B^t|_{[1]}\nu_a^t \wedge \pi) \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_a^t|\pi)} \\ \times \sum_{S^t} \left[\begin{array}{l} P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\ \times \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \times P(S^t|\nu_p^t \wedge \pi) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_a^t|\pi)} \times \kappa_{t-1} \times \frac{P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi)}{\prod_{i=1}^{N_i}\left[P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i)\right]} \\ \times \sum_{S^t} \left[\begin{array}{l} \prod_{i=1}^{N_i}\left[P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}\beta_i^t|_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i)\right] \\ \times \prod_{i=1}^{N_i}\left[P({}_{[1]}C^t|\pi_i) \times P({}_{[1]}\alpha_i^t|_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i)\right] \\ \times \prod_{i=1}^{N_i} P(S_i^t|\nu_p^t \wedge \pi_i) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_a^t|\pi)} \times \kappa_{t-1} \times \frac{P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi)}{\prod_{i=1}^{N_i}\left[P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i)\right]} \\ \times \sum_{S^t} \left[\prod_{i=1}^{N_i} \left[\begin{array}{l} P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}\beta_i^t|_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \times P({}_{[1]}C^t|\pi_i) \times P({}_{[1]}\alpha_i^t|_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \times P(S_i^t|\nu_p^t\pi_i) \end{array} \right] \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_a^t|\pi)} \times \kappa_{t-1} \times \frac{P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi)}{\prod_{i=1}^{N_i}\left[P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i)\right]} \\ \prod_{i=1}^{N_i} \left[\sum_{S_i^t} \left[\begin{array}{l} P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}\beta_i^t|_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}B^{t-1}\pi_i) \\ \times P({}_{[1]}C^t|\pi_i) \times P({}_{[1]}\alpha_i^t|_{[1]}C^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \times P(S_i^1|\nu_p^t \wedge \pi_i) \end{array} \right] \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_a^t|\pi)} \times \kappa_{t-1} \times \frac{P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi)}{\prod_{i=1}^{N_i}\left[P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i)\right]} \prod_{i=1}^{N_i}\left[\frac{P({}_{[t]}\nu_a^{\downarrow}\pi_i)}{P(\nu_p^t|\pi_i)}\right] \\ \prod_{i=1}^{N_i}\left[\ P({}_{[1]}C^t \wedge {}_{[1]}B^t|_{[1]}\nu_a^t \wedge \pi_i)\ \right] \end{array} \right.
\end{aligned}$$

Now, we desire to proof that this proportionality happens for all other levels in the hierarchy.

Right part of attention Selection question at level h, so that $2 \leq h \leq L$ is:

$$
{}_{[h]}\nu_a^t = \left| \begin{array}{l} {}_{[1:h-1]}c^{0:t} \wedge {}_{[1:h]}\alpha^{0:t} \wedge {}_{[h:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha^{0:t-1} \wedge \\ {}_{[1:h-1]}b^{0:t} \wedge {}_{[1:h]}\beta^{0:t} \wedge {}_{[h:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta^{0:t-1} \wedge \\ {}_{[1:h-1]}z^{0:t} \wedge {}_{[h:L]}z^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \end{array} \right.
$$

$$
\begin{aligned}
&P({}_{[h]}C^t \wedge {}_{[h]}B^t | {}_{[h]}\nu_a^t \wedge \pi) \\
&\quad = P({}_{[h]}C^t \wedge {}_{[h]}B^t | \left[\begin{array}{l} {}_{[1:h-1]}c^{0:t} \wedge {}_{[1:h]}\alpha^{0:t} \wedge {}_{[h:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha^{0:t-1} \wedge \\ {}_{[1:h-1]}b^{0:t} \wedge {}_{[1:h]}\beta^{0:t} \wedge {}_{[h:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta^{0:t-1} \wedge \\ {}_{[1:h-1]}z^{0:t} \wedge {}_{[h:L]}z^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi \end{array} \right])
\end{aligned}
$$

And developing the question for attention Selection question at level h, so that $2 \leq h \leq L$, we have:

$$
\begin{aligned}
&P({}_{[h]}C^t \wedge {}_{[h]}B^t | {}_{[h]}\nu_a^t \wedge \pi) \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_a^t|\pi)} \\ \times \sum_{S^t} \left[\begin{array}{l} P({}_{[1]}b^t|\pi) \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\ \times P({}_{[1]}c^t|\pi) \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[1]}z_i^t | S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{k=2}^{h-1} \left[\begin{array}{l} P({}_{[k]}b^t|\pi) \ \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\ \times P({}_{[k]}c^t|\pi) \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\ \prod_{i=1}^{N_i} P({}_{[k]}z_i^t | S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array} \right] \\ \times P({}_{[h]}b^t|\pi) \prod_{i=1}^{N_i} P({}_{[h]}\beta_i^t|{}_{[h]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[h]}b^{t-1}\pi_i) \\ \times P({}_{[h]}c^t|\pi) \prod_{i=1}^{N_i} P({}_{[h]}\alpha_i^t|{}_{[h]}c^t \wedge S_i^t \wedge {}_{[h]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\ \times P(S^t|\nu_p^t \wedge \pi) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_a^t|\pi)} \prod_{k=1}^{h} \left[\frac{P({}_{[k]}b^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} \left[P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \right]} \right] \times \kappa_{t-1} \\ \times \sum_{S^t} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \left[P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \right] \\ \prod_{i=1}^{N_i} P({}_{[1]}z_i^t | S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{k=2}^{h-1} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \left[P({}_{[k]}c^t|\pi_i) \times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \right] \\ \prod_{i=1}^{N_i} P({}_{[k]}z_i^t | S_i^t \wedge {}_{[k]}C^t \wedge \pi_i) \end{array} \right] \\ \prod_{i=1}^{N_i} \left[P({}_{[h]}b^t|\pi_i) \times P({}_{[h]}\beta_i^t|{}_{[h]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[h]}b^{t-1}\pi_i) \right] \\ \prod_{i=1}^{N_i} \left[P({}_{[h]}c^t|\pi_i) \times P({}_{[h]}\alpha_i^t|{}_{[h]}c^t \wedge S_i^t \wedge {}_{[h]}b^t \wedge {}_{[k-1]}b^t\pi_i) \right] \\ \prod_{i=1}^{N_i} P(S_i^t|\nu_p^t \wedge \pi_i) \end{array} \right] \end{array} \right.
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_a^t|\pi)} \prod_{k=1}^{h} \left[\frac{P({}_{[k]}b^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} \left[P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \right]} \right] \times \kappa_{t-1} \\
\times \sum_{S^t} \left[\prod_{i=1}^{N_i} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h-1} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}c^t|\pi_i) \times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P({}_{[h]}b^t|\pi_i) \times P({}_{[h]}\beta_i^t|{}_{[h]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\times P({}_{[h]}c^t|\pi_i) \times P({}_{[h]}\alpha_i^t|{}_{[h]}c^t \wedge S_i^t \wedge {}_{[h]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array} \right] \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_a^t|\pi)} \prod_{k=1}^{h} \left[\frac{P({}_{[k]}b^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} \left[P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \right]} \right] \times \kappa_{t-1} \\
\prod_{i=1}^{N_i} \left[\sum_{S_i^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h-1} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}c^t|\pi_i) \times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P({}_{[h]}b^t|\pi_i) \times P({}_{[h]}\beta_i^t|{}_{[h]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[h]}b^{t-1}\pi_i) \\
\times P({}_{[h]}c^t|\pi_i) \times P({}_{[h]}\alpha_i^t|{}_{[h]}c^t \wedge S_i^t \wedge {}_{[h]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array} \right] \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_a^t|\pi)} \prod_{k=1}^{h} \left[\frac{P({}_{[k]}b^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} \left[P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \right]} \right] \\
\times \kappa_{t-1} \prod_{i=1}^{N_i} \left[\frac{P({}_{[t]}\nu_a \pi_i)}{P(\nu_p^t|\pi_i)} \right] \\
\prod_{i=1}^{N_i} \left[\; P({}_{[h]}C^t \wedge {}_{[h]}B^t|{}_{[h]}\nu_a^t \wedge \pi_i) \; \right]
\end{array} \right|
$$

Behaviour Selection

It is desired to show behaviour selection question at level 1 to Global distribution is proportional to the product of behaviour selection question at level 1 in the elementary filters.

Right part of behaviour selection at level 1 for Global filter is:

$$
{}_{[1]}\nu_b^t = \left| \begin{array}{l}
{}_{[1]}c^{0:t} \wedge {}_{[1]}\alpha^{0:t} \wedge {}_{[2:L]}c^{0:t-1} \wedge {}_{[2:L]}\alpha^{0:t-1} \wedge \\
{}_{[1]}z^{0:t} \wedge {}_{[2:L]}z^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1]}\beta^{0:t} \wedge {}_{[2:L]}\beta^{0:t-1} \wedge \\
m^{0:t-1} \wedge \lambda^{0:t-1}
\end{array} \right.
$$

$$
\begin{aligned}
& P({}_{[1]}B^t|{}_{[1]}\nu_b^t \wedge \pi) \\
& \quad = P({}_{[1]}B^t| \left[\begin{array}{l}
{}_{[1]}c^{0:t} \wedge {}_{[1]}\alpha^{0:t} \wedge {}_{[2:L]}c^{0:t-1} \wedge {}_{[2:L]}\alpha^{0:t-1} \wedge \\
{}_{[1]}z^{0:t} \wedge {}_{[2:L]}z^{0:t-1} \wedge {}_{[1:L]}b^{0:t-1} \wedge {}_{[1]}\beta^{0:t} \wedge {}_{[2:L]}\beta^{0:t-1} \wedge \\
m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi
\end{array} \right])
\end{aligned}
$$

Developing question for Behaviour Selection question at level 1 for global filter, we find:

$$
\begin{aligned}
&P({}_{[1]}B^t|_{[1]}\nu_b^t \wedge \pi) \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_b^t|\pi)} \\ \times \sum_{S^t} \left[\begin{array}{l} P({}_{[1]}B^t|\pi) \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}B^{t-1}\pi_i) \\ \times P({}_{[1]}c^t|\pi) \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \times \prod_{i=1}^{N_i} P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ P(S^t|\nu_p^t \wedge \pi) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_b^t|\pi)} \times \kappa_{t-1} \frac{P({}_{[1]}B^t|\pi) \times P({}_{[1]}c^t|\pi)}{\prod_{i=1}^{N_i} \left[P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}c^t|\pi_i)\right]} \\ \times \sum_{S^t} \left[\begin{array}{l} \prod_{i=1}^{N_i} \left[P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}\beta_i^t|_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}B^{t-1}\pi_i)\right] \\ \prod_{i=1}^{N_i} \left[P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i)\right] \\ \prod_{i=1}^{N_i} P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \prod_{i=1}^{N_i} P(S_i^t|\nu_p^t \wedge \pi_i) \end{array} \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_b^t|\pi)} \times \kappa_{t-1} \frac{P({}_{[1]}B^t|\pi) \times P({}_{[1]}c^t|\pi)}{\prod_{i=1}^{N_i} \left[P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}c^t|\pi_i)\right]} \\ \prod_{i=1}^{N_i} \left[\sum_{S_i^t} \left[\begin{array}{l} P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}\beta_i^t|_{[1]}B^t \wedge S_i^t \wedge {}_{[1]}B^{t-1}\pi_i) \\ \times P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}B^t\pi_i) \\ \times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\ \times P(S_i^t|\nu_p^t \wedge \pi_i) \end{array} \right] \right] \end{array} \right. \\
&= \left| \begin{array}{l} \frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_b^t|\pi)} \times \kappa_{t-1} \frac{P({}_{[1]}B^t|\pi) \times P({}_{[1]}c^t|\pi)}{\prod_{i=1}^{N_i} \left[P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}c^t|\pi_i)\right]} \prod_{i=1}^{N_i} \left[\frac{P({}_{[t]}\nu_b^{\downarrow}\pi_i)}{P(\nu_p^t|\pi_i)}\right] \\ \prod_{i=1}^{N_i} \left[\ P({}_{[1]}B^t|_{[1]}\nu_b^t \wedge \pi_i)\ \right] \end{array} \right.
\end{aligned}
$$

Now, we desire to proof that this proportionality happens for all other levels in the hierarchy.

Right part of Behaviour Selection question at level h, so that $2 \leq h \leq L$ is:

$$
{}_{[h]}\nu_b^t = \left| \begin{array}{l} {}_{[1:h]}c^{0:t} \wedge {}_{[1:h]}\alpha^{0:t} \wedge {}_{[h+1:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha^{0:t-1} \wedge \\ {}_{[1:h-1]}b^{0:t} \wedge {}_{[1:h]}\beta^{0:t} \wedge {}_{[h+1:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta^{0:t-1} \wedge \\ {}_{[1:h]}z^{0:t} \wedge {}_{[h=1:L]}z^{0:t-1} \wedge m^{0:t-1} \wedge \lambda_i^{0:t-1} \end{array} \right.
$$

$$
\begin{aligned}
&P({}_{[h]}B^t|_{[h]}\nu_b^t \wedge \pi) \\
&\quad = P({}_{[h]}B^t \left[\begin{array}{l} {}_{[1:h]}c^{0:t} \wedge {}_{[1:h]}\alpha^{0:t} \wedge {}_{[h+1:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha^{0:t-1} \wedge \\ {}_{[1:h-1]}b^{0:t} \wedge {}_{[1:h]}\beta^{0:t} \wedge {}_{[h+1:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta^{0:t-1} \wedge \\ {}_{[1:h]}z^{0:t} \wedge {}_{[h=1:L]}z^{0:t-1} \wedge m^{0:t-1} \wedge \lambda_i^{0:t-1} \wedge \pi \end{array} \right])
\end{aligned}
$$

And developing the question for Behaviour Selection question at level h, so that $2 \leq h \leq L$, we have:

$$
\begin{aligned}
&P({}_{[h]}B^t|_{[h]}\nu_b^t \wedge \pi) \\
&= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_b^t|\pi)} \\
\times \sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi) \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}c^t|\pi) \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}B^{t-1}\pi_i) \\
\times P({}_{[k]}c^t|\pi) \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i) \\
\times
\end{array} \right] \\
\times P(S^t|\nu_p^t \wedge \pi)
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_b^t|\pi)} \times \kappa_{t-1} \prod_{k=1}^{h} \left[\frac{P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} \left[P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \right]} \right] \\
\times \sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
\prod_{i=1}^{N_i} P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}B^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}c^t|\pi_i) \times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i) \\
\times
\end{array} \right] \\
\prod_{i=1}^{N_i} \left[\; P(S_i^t|\nu_p^t \wedge \pi) \; \right]
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_b^t|\pi)} \times \kappa_{t-1} \prod_{k=1}^{h} \left[\frac{P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} \left[P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \right]} \right] \\
\times \sum_{S^t} \left[\prod_{i=1}^{N_i} \left[\begin{array}{l}
P(S_i^t \,|\, S_i^{t-1} \wedge m^{t-1} \wedge \pi_i) \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}B^{t-1}\pi_i) \\
\times P({}_{[k]}c^t|\pi_i) \times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i) \\
\times
\end{array} \right] \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array} \right] \right]
\end{array} \right|
\end{aligned}
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_b^t|\pi)} \times \kappa_{t-1} \times \frac{P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi)}{\prod} {}^h_{k=1} \left[\frac{P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} \left[P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \right]} \right] \\
\times \prod_{i=1}^{N_i} \left[\sum_{S_i^t} \left[\begin{array}{l}
P(S_i^t \left| S_i^{t-1} \wedge m^{t-1} \wedge \pi_i \right.) \\
\times P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}\beta_i^t|_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}B^t \wedge {}_{[k]}B^{t-1}\pi_i) \\
\times P({}_{[k]}c^t|\pi_i) \times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}z_i^t \,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i) \\
\times
\end{array} \right] \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array} \right] \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_b^t|\pi)} \times \kappa_{t-1} \prod_{k=1}^{h} \left[\frac{P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} \left[P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \right]} \right] \prod_{i=1}^{N_i} \left[\frac{P({}_{[t]}\nu_b^{|}\pi_i)}{P(\nu_p^t|\pi_i)} \right] \\
\prod_{i=1}^{N_i} \left[\; P({}_{[h]}B^t|_{[h]}\nu_b^t \wedge \pi_i) \; \right]
\end{array} \right|
$$

Estimation

It is desired to show that estimation question at level 1 to Global distribution is proportional to the product of estimation question at level 1 in the elementary filters.

Right part of estimation question at level 1 for Global filter is:

$$
{}_{[1]}\nu_e^t = \left[\begin{array}{l}
{}_{[1]}c^{0:t} \wedge {}_{[1]}\alpha^{0:t} \wedge {}_{[2:L]}c^{0:t-1} \wedge {}_{[2:L]}\alpha^{0:t-1} \wedge \\
{}_{[1]}z^{0:t} \wedge {}_{[2:L]}z^{0:t-1} \wedge {}_{[1]}b^{0:t} \wedge {}_{[1]}\beta^{0:t} \wedge \\
{}_{[2:L]}b^{0:t-1} \wedge {}_{[2:L]}\beta^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi
\end{array} \right]
$$

$$
\begin{array}{l}
P(S^t|_{[1]}\nu_e^t \wedge \pi) \\
\quad = P(S^t| \left[\begin{array}{l}
{}_{[1]}c^{0:t} \wedge {}_{[1]}\alpha^{0:t} \wedge {}_{[2:L]}c^{0:t-1} \wedge {}_{[2:L]}\alpha^{0:t-1} \wedge \\
{}_{[1]}z^{0:t} \wedge {}_{[2:L]}z^{0:t-1} \wedge {}_{[1]}b^{0:t} \wedge {}_{[1]}\beta^{0:t} \wedge \\
{}_{[2:L]}b^{0:t-1} \wedge {}_{[2:L]}\beta^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi
\end{array} \right])
\end{array}
$$

Developing estimation question expression at level 1 for global filter, we find:

$$
\begin{array}{l}
P(S^t|_{[1]}\nu_e^t \wedge \pi) \\
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_e^t|\pi)} \\
\times P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\times P(S^t|\nu_p^t \wedge \pi)
\end{array} \right.
\end{array}
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_e^t|\pi)} \times \frac{P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi)}{\prod_{i=1}^{N_i} P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i)} \\
\prod_{i=1}^{N_i} \{P(P({}_{[1]}B^t|\pi_i) \times {}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i)\} \\
\prod_{i=1}^{N_i} \{P(P({}_{[1]}C^t|\pi_i) \times {}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i)\} \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\times \kappa_{t-1} \prod_{i=1}^{N_i} P(S^t|\nu_p^t \wedge \pi_i)
\end{array} \right.
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_e^t|\pi)} \times \kappa_{t-1} \times \frac{P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi)}{\prod_{i=1}^{N_i} \{P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i)\}} \\
\prod_{i=1}^{N_i} \left[\begin{array}{l}
P(P({}_{[1]}B^t|\pi_i) \times {}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P(P({}_{[1]}C^t|\pi_i) \times {}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\times P(S^t|\nu_p^t \wedge \pi_i)
\end{array} \right]
\end{array} \right.
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[1]}\nu_e^t|\pi)} \times \kappa_{t-1} \times \prod_{i=1}^{N_i} \frac{P({}_{[t]}\nu_e^{\downarrow}\pi_i)}{P(\nu_p^t|\pi_i)} \\
\times \frac{P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi)}{\prod_{i=1}^{N_i} \{P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i)\}} \\
\prod_{i=1}^{N_i} P(S^t|{}_{[1]}\nu_e^t \wedge \pi_i)
\end{array} \right.
$$

Now, we desire to proof that this proportionality happens for all other levels in the hierarchy.

Right part of estimation question at level h, so that $2 \leq h \leq L$ is:

$$
{}_{[h]}\nu_e^t = \left[\begin{array}{l}
{}_{[1:h]}c^{0:t} \wedge {}_{[1:h]}\alpha^{0:t} \wedge {}_{[h+1:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha^{0:t-1} \wedge \\
{}_{[1:h]}b^{0:t} \wedge {}_{[1:h]}\beta^{0:t} \wedge {}_{[h+1:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta^{0:t-1} \wedge \\
{}_{[1:h]}z^{0:t} \wedge {}_{[h+1:L]}z^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1}
\end{array} \right]
$$

$$
\begin{array}{l}
P(S^t|{}_{[h]}\nu_e^t \wedge \pi) \\
\quad = P(S^t| \left[\begin{array}{l}
{}_{[1:h]}c^{0:t} \wedge {}_{[1:h]}\alpha^{0:t} \wedge {}_{[h+1:L]}c^{0:t-1} \wedge {}_{[h+1:L]}\alpha^{0:t-1} \wedge \\
{}_{[1:h]}b^{0:t} \wedge {}_{[1:h]}\beta^{0:t} \wedge {}_{[h+1:L]}b^{0:t-1} \wedge {}_{[h+1:L]}\beta^{0:t-1} \wedge \\
{}_{[1:h]}z^{0:t} \wedge {}_{[h+1:L]}z^{0:t-1} \wedge m^{0:t-1} \wedge \lambda^{0:t-1} \wedge \pi
\end{array} \right])
\end{array}
$$

And developing estimation question expression at level h, so that $2 \leq h \leq L$, we have:

$$
P(S^t|{}_{[h]}\nu_e^t \wedge \pi)
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_e^t|\pi)} \\
\times P({}_{[1]}b^t|\pi) \times P({}_{[1]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi) \\
\prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P(S^t|\nu_p^t \wedge \pi)
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_e^t|\pi)} \times \frac{P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi)}{\prod_{i=1}^{N_i} P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i)} \prod_{k=2}^{h} \left\{ \frac{P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} [P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i)]} \right\} \\
\prod_{i=1}^{N_i} [P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i)] \\
\prod_{i=1}^{N_i} [P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i)] \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
\prod_{i=1}^{N_i} [P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i)] \\
\prod_{i=1}^{N_i} [P({}_{[k]}c^t|\pi_i) \times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i)] \\
\prod_{i=1}^{N_i} P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times \kappa_{t-1} \prod_{i=1}^{N_i} P(S^t|\nu_p^t \wedge \pi_i)
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_e^t|\pi)} \times \frac{P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi)}{\prod_{i=1}^{N_i} P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i)} \prod_{k=2}^{h} \left\{ \frac{P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} [P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i)]} \right\} \\
\times \kappa_{t-1} \\
\prod_{i=1}^{N_i} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
P({}_{[1]}c^t|\pi) \times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{h} \left[\begin{array}{l}
P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i) \\
\times P({}_{[k]}\beta_i^t|{}_{[k]}B^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}B^t \wedge {}_{[k-1]}b^t\pi_i) \\
P({}_{[k]}z_i^t\,|S_i^t \wedge {}_{[k]}c^t \wedge \pi_i)
\end{array} \right] \\
\times P(S^t|\nu_p^t \wedge \pi_i)
\end{array} \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P({}_{[h]}\nu_e^t|\pi)} \times \frac{P({}_{[1]}B^t|\pi) \times P({}_{[1]}C^t|\pi)}{\prod_{i=1}^{N_i} P({}_{[1]}B^t|\pi_i) \times P({}_{[1]}C^t|\pi_i)} \prod_{k=2}^{h} \left\{ \frac{P({}_{[k]}B^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} [P({}_{[k]}B^t|\pi_i) \times P({}_{[k]}c^t|\pi_i)]} \right\} \\
\times \kappa_{t-1} \prod_{i=1}^{N_i} \frac{P({}_{[h]}\nu_e^t|\pi_i)}{P(\nu_p^t|\pi_i)} \\
\prod_{i=1}^{N_i} P(S^t|{}_{[h]}\nu_e^t \wedge \pi)
\end{array} \right|
$$

Motor Command

It is desired to show motor command question to Global distribution is proportional to the product of motor command questions to elementary filters.

Right part of motor command question is:

$$\nu_m^t = \Big| \; {}_{[1:L]}c^{0:t} \wedge {}_{[1:L]}\alpha^{0:t} \wedge {}_{[1:L]}z^{0:t} \wedge {}_{[1:L]}b^{0:t} \wedge {}_{[1:L]}\beta^{0:t} \wedge m^{0:t-1} \wedge \lambda^{0:t}$$

$$\begin{aligned}
&P(M^t|\nu_m^t \wedge \pi) \\
&\quad = P(M^t| \left[\; {}_{[1:L]}c^{0:t} \wedge {}_{[1:L]}\alpha^{0:t} \wedge {}_{[1:L]}z^{0:t} \wedge {}_{[1:L]}b^{0:t} \wedge {}_{[1:L]}\beta^{0:t} \wedge m^{0:t-1} \wedge \lambda^{0:t} \wedge \pi \; \right])
\end{aligned}$$

And motor command question to Global distribution is:

$$\begin{aligned}
&P(M^t|\nu_m^t \wedge \pi) \\
&= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P(\nu_m^t|\pi)} \\
\times \sum_{S^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi) \prod_{i=1}^{N_i} P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}c^t|\pi) \prod_{i=1}^{N_i} P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi) \prod_{i=1}^{N_i} P({}_{[k]}\beta_i^t|_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}c^t|\pi) \prod_{i=1}^{N_i} P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi) \prod_{i=1}^{N_i} P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}b^t \wedge S_i^t\pi_i) \\
\times P(S^t|\nu_m^t \wedge \pi)
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P(\nu_m^t|\pi)} \times \kappa_{t-1} \times \frac{P(M^t|\pi)}{\prod_{i=1}^{N_i}[P(M^t|\pi_i)]} \prod_{k=1}^{L} \left[\frac{P({}_{[k]}b^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i}\left[P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i)\right]} \right] \\
\times \sum_{S^t} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \right] \\
\prod_{i=1}^{N_i} \left[P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \right] \\
\prod_{i=1}^{N_i} P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
\prod_{i=1}^{N_i} \left[P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}\beta_i^t|_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \right] \\
\prod_{i=1}^{N_i} \left[P({}_{[k]}c^t|\pi_i) \times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \right] \\
\prod_{i=1}^{N_i} P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\prod_{i=1}^{N_i} \left[P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}b^t \wedge S_i^t\pi_i) \right] \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array} \right]
\end{array} \right| \\
&= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P(\nu_m^t|\pi)} \times \kappa_{t-1} \times \frac{P(M^t|\pi)}{\prod_{i=1}^{N_i} P(M^t|\pi_i)} \\
\prod_{k=1}^{L} \frac{P({}_{[k]}b^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i)} \\
\times \sum_{S^t} \left[\prod_{i=1}^{N_i} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t \,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}\beta_i^t|_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}c^t|\pi_i) \times P({}_{[k]}\alpha_i^t|_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}Z_i^t \,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}b^t \wedge S_i^t\pi_i) \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array} \right] \right]
\end{array} \right|
\end{aligned}$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P(\nu_m^t|\pi)} \times \kappa_{t-1} \times \frac{P(M^t|\pi)}{\prod_{i=1}^{N_i} P(M^t|\pi_i)} \prod_{k=1}^{L} \frac{P({}_{[k]}b^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i)} \\
\prod_{i=1}^{N_i} \left[\sum_{S_i^t} \left[\begin{array}{l}
P({}_{[1]}b^t|\pi_i) \times P({}_{[1]}\beta_i^t|{}_{[1]}b^t \wedge S_i^t \wedge {}_{[1]}b^{t-1}\pi_i) \\
\times P({}_{[1]}c^t|\pi_i) \times P({}_{[1]}\alpha_i^t|{}_{[1]}c^t \wedge S_i^t \wedge {}_{[1]}b^t\pi_i) \\
\times P({}_{[1]}z_i^t\,|S_i^t \wedge {}_{[1]}c^t \wedge \pi_i) \\
\prod_{k=2}^{L} \left[\begin{array}{l}
P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}\beta_i^t|{}_{[k]}b^t \wedge S_i^t \wedge {}_{[k-1]}b^t \wedge {}_{[k]}b^{t-1}\pi_i) \\
\times P({}_{[k]}c^t|\pi_i) \times P({}_{[k]}\alpha_i^t|{}_{[k]}c^t \wedge S_i^t \wedge {}_{[k]}b^t \wedge {}_{[k-1]}b^t\pi_i) \\
\times P({}_{[k]}Z_i^t\,|S_i^t \wedge {}_{[k]}C^t \wedge \pi_i)
\end{array} \right] \\
\times P(M^t|\pi_i) \times P(\lambda_i^t|M^t \wedge m^{t-1} \wedge {}_{[L]}b^t \wedge S_i^t\pi_i) \\
\times P(S_i^t|\nu_p^t \wedge \pi_i)
\end{array} \right] \right]
\end{array} \right|
$$

$$
= \left| \begin{array}{l}
\frac{P(\nu_p^t|\pi)}{P(\nu_m^t|\pi)} \times \kappa_{t-1} \\
\times \frac{P(M^t|\pi)}{\prod_{i=1}^{N_i} P(M^t|\pi_i)} \\
\prod_{k=1}^{L} \frac{P({}_{[k]}b^t|\pi) \times P({}_{[k]}c^t|\pi)}{\prod_{i=1}^{N_i} P({}_{[k]}b^t|\pi_i) \times P({}_{[k]}c^t|\pi_i)} \prod_{i=1}^{N_i} \frac{P(\nu_m^t|\pi_i)}{P(\nu_p^t|\pi_i)} \\
\prod_{i=1}^{N_i} \left[\; P(M^t|\nu_m^t \wedge \pi_i) \; \right]
\end{array} \right|
$$

E.4 Requirements Analysis

Requirements analysis for this proposal is restricted to complexity expressions.

E.4.1 Space Complexity

- Dynamic model: $n^{2N_{s_i}+N_m}$
- Sensor model at level h, $1 \leq h \leq L$:
  $n^{{}_{[h]}N_{z_i}+N_{s_i}+{}_{[h]}N_c)}$
- Attention selection model at level h, $1 \leq h \leq L$:
  $n^{2+{}_{[h]}N_c+2{}_{[h]}N_b+N_{s_i}}$
- Behaviour selection model at level h, $1 \leq h \leq L$:
  $n^{2+3{}_{[h]}N_b+N_{s_i}}$
- Motor model:$n^{2+2N_m+{}_{[L]}N_b+N_{s_i}}$

Each question needs also to be stored: one prediction ($n^{N_{s_i}}$) and one motor question (n^{N_m}) for each elementary filter i. For each level h in each elementary filter i, there are also one attention question ($n^{{}_{[h]}N_c}$), one behaviour question $n^{{}_{[h]}N_b}$ and one estimation question ($n^{N_{s_i}}$). Additionally, one fusion result question for each level: $L \times n^{{}_{[h]}N_c}$, for attention, and $L \times n^{{}_{[h]}N_b}$, for behaviour.

$$\mathcal{M} = \left| \begin{array}{l} \sum_{i=1}^{N_i} \left[\begin{array}{l} n^{2N_{s_i}+N_m} + n^{2+2N_m+{}_{[L]}N_b+N_{s_i}} \\ + \sum_{h=1}^{L} \left[\begin{array}{l} n^{({}_{[h]}N_{z_i} N_{s_i} + {}_{[h]}N_c)} \\ +n^{2+{}_{[h]}N_c+2{}_{[h]}N_b+N_{s_i}} \\ +n^{2+3{}_{[h]}N_b+N_{s_i}} \\ +n^{{}_{[h]}N_c} + n^{{}_{[h]}N_b} \end{array} \right] \\ +(L+1)n^{N_{s_i}} \end{array} \right] \\ + \sum_{h=1}^{L} \left[n^{{}_{[h]}N_c} + n^{{}_{[h]}N_b} \right] + N_i \times n^{N_m} \end{array} \right|$$

E.4.2 Time Complexity

- Prediction $n^{2N_{s_i}} - n^{N_{s_i}}$
- Attention question for each level h, $1 \leq h \leq L$:
  $n^{({}_{[h]}N_c+{}_{[h]}N_b)}[3n^{N_{s_i}} - 1]$
- Attention fusion for each level h, $1 \leq h \leq L$:
  $N_i \times n^{({}_{[h]}N_c+{}_{[h]}N_b)}$
- Behaviour question for each level h, $1 \leq h \leq L$:
  $n^{{}_{[h]}N_b} \times (4n^{N_{s_i}} - 1)$
- Behaviour fusion for each level h, $1 \leq h \leq L$:
  $N_i \times n^{{}_{[h]}N_b}$
- Estimation for each level h, $1 \leq h \leq L$:
  $3n^{N_{s_i}}$
- Motor question and fusion: $2n^{N_{s_i}+N_m}$

Summing all terms above:

$$\mathcal{C} = \sum_{i=1}^{N_i} \left\{ \begin{array}{l} n^{2N_{s_i}} - n^{N_{s_i}} + 2n^{N_{s_i}+N_m} \\ \sum_{h=1}^{L} \left[\begin{array}{l} n^{({}_{[h]}N_c+{}_{[h]}N_b)}[3n^{N_{s_i}} - 1] \\ +n^{{}_{[h]}N_b} \times (4n^{N_{s_i}} - 1) \\ +3n^{N_{s_i}} \end{array} \right] \end{array} \right\} + \sum_{h=1}^{L} \left[\begin{array}{l} N_i \times n^{({}_{[h]}N_c+{}_{[h]}N_b)} \\ +N_i \times n^{{}_{[h]}N_b} \end{array} \right]$$

Part II

Implementation Details

Appendix F

Simulation Examples

Probability tables for behaviour and attention models used by simulation examples are only briefly described in chapters 6 and 7. This appendix presents the complete table description for behaviour selection model introduced in chapter 6, and for Attention selection model introduced in chapter 7. Furthermore, implementation source code is also available in www.laplace.imag.fr for inspection.

F.1 Behaviour Selection Model

Chapter 6 introduced the variable behaviour in our illustration example. Also, the behaviour selection model in introduced in the joint distribution of all elementary filters, and this model is programmed as probability tables. Predator filter table was detailled and discussed in chapter 6(table 6.1) , and prey filter table was presented in chapter 7(table 7.1).

Home filter behaviour selection model is described here.

H^j_{dist}	0,1				2				3,4			
B^{j-1} / B^j	E	C	H	M	E	C	H	M	E	C	H	M
E	+	∘	∘	∘	+	∘	∘	∘	+	∘	∘	∘
C	∘	+	∘	∘	∘	+	∘	∘	∘	+	∘	∘
H	++	+	++	++	+	∘	+	+	∘	∘	+	∘
M	∘	∘	∘	+	∘	∘	∘	+	∘	∘	∘	+

Table F.1: Probability table for $P(B^j|B^{j-1}H^j_{dist}\pi_{home})$, behaviour selection direct model for home filter. E stands for Escape behaviour; C denotes Chase behaviour; H indicates Homing behaviour; and M signifies Motionless behaviour. ++ means highly recommended, + means recommended, ∘ means unconcerned, - means unwise and × stands for forbidden. An asterisk is employed when situation is the same for all values of one variable.

F.2 Attention Selection Model

Attention model was proposed in chapter 7, and table for attention selection model in prey filter was described in table 7.2.

F.2.1 Predator Filter

X^j_{dist}	0,1,2,3				3,4			
C^j \ B^t	E	C	H	M	E	C	H	M
Predator	++	++	+	+	++	+	+	++
Prey	×	×	×	×	∘	+	∘	∘
Home	++	++	++	++	+	∘	++	∘

Table F.2: Probability table for $P(C^j|X^j_{dist}B^t\pi_{predator})$, attention selection direct model for predator filter. E stands for Escape behaviour; C denotes Chase behaviour; H indicates Homing behaviour; and M signifies Motionless behaviour. ++ means highly recommended, + means recommended, and ∘ means unconcerned.

F.2.2 Home Filter

H^j_{dist}	0,1,2				3,4			
C^j \ B^t	E	C	H	M	E	C	H	M
Predator	++	+	+	+	++	+	+	++
Prey	×	+	×	∘	∘	++	∘	∘
Home	+	∘	++	++	+	∘	++	∘

Table F.3: Probability table for $P(C^j|H^j_{dist}B^t\pi_{home})$, attention selection direct model for home filter. E stands for Escape behaviour; C denotes Chase behaviour; H indicates Homing behaviour; and M signifies Motionless behaviour. ++ means highly recommended, + means recommended, and ∘ means unconcerned.

Appendix G

Image Processing Algorithm

G.1 Requirements

Used operating systems is Linux, kernel version 2.6, with Eclipse 3.1 and compiler gcc version 3.xx.

Two libraries were required for dealing with images. blinky v1.1.1 is necessary for managing image acquisition from a FireWire camera connexion. This library was developed by MOVI team (www.inrialpes.fr/movi, contact Frederic Devernay).

Another library used is OpenCV v0.9.6, for image processing and visualisation. The sources can be found in http://sourceforge.net/projects/opencvlibrary/.

Two camera were tested, both of them connected via FireWire 1394: IBot camera (http://www.firewire-1394.com/orange-micro-ibot-firewire-desktop-video-camera.htm) and Avt Marlin (http://www.imasys.fr/products/FR/643.htm).

All programs were developed in C and C++ languages.

G.2 Vision Processing

It is desired to detect the zone in the image where a particular colour is located.

A range of values for hue and saturation is chosen characterising each colour. This range of values are used to build a two dimension histogram. It is necessary one histogram for each different colour.

G.2.1 Detection Algorithm

1. Image is acquired with two possible resolutions: 640x480 or 320x240.

2. Back-projection of the captured image is calculated employing the two dimension histogram (hue and saturation) of this image and the histogram of searched colour. Result is a black and white image where white zones correspond to zones in the original image of the searched colour.

3. Resulting black and white image is thresholded in order to delete noise pixels.
4. Search of white zones contours in black and white image.
5. Biggest zone is selected as dominant. A minimal surface area is defined for a contour to be considered as valid. Also, a minimal ratio is also predefined: minimal surface is 2.0/200.0 and ration must be greater than 0.5
6. A rectangular contour bounding the selected zone is calculated.
7. Position of the zone is defined as the center of gravity of the zone.

Appendix H

Robot Experiments

The presentation will be based on the elementary filters, and there are four ones in this experiment: prey, trainer, predator and escape route. For each filter, presentation will follow Bayesian Robot Programming general structure: relevant variables, joint distribution decomposition, parametric forms and identification, and utilisation.

H.1 Types Description

Presence

Binary integer type: value zero means not present, while value one means present.

Distance

Integer type, whose values are in the range 0 to 5. 0 means very near, and 5 means far away. These values are related to size of area detected in the image. It is not a precise measure of distance.

Direction

Integer type indicating zone surrounding robot where an object can be located: from -4 to 4, according to illustration in figure 8.6.

Behaviour

Behaviour variable is an integer, with five possible values, one for each possible motor pattern: *Escape*, *Motionless*, *Obey*, *Chase* and *Wander*.

Translation Speed

Integer value indicating how fast robot advances in its forward direction. Four possible values, where 0 means the robot does not move forward, and 3 means it progresses as fastest as possible. Biba robot can achieve an speed of 3 meters per second, but for security measures, maximum speed is limited to 1 meter per second.

Rotation Speed

Rotation speed variable is of integer type, with 9 values: -4 means full rotation speed turning left and 4 means full rotation speed turning right.

Coherence based Fusion

Coherence variables employed in this fusion schema are always of binary integer type: zero value means not coherence, while value one means coherence.

H.1.1 Common Variables

Motor commands are common to all filters and they are here rotation speed M_{rot} (nine values) and translation speed M_{trans} (six values). Figure 8.7 shows how translation and rotation speed and related to the robot.

Behaviour variable B^t is also common to all elementary filter, with values *Escape*, *Motionless*, *Obey*, *Chase* and *Wander*. Escape, motionless and chase behaviour have the same meaning employed in the simulation example, as explained in example 2 of section 4.1.3 and in section 6.1.3.

Obey behaviour means following trainer at a near distance, as wander behaviour denotes that robot progress at medium translation speed in the forward direction, avoiding obstacles.

Head pan direction is an observation variable ZD_{pan} used by all filters, and it is employed to verify if object in inside robot field of view.

H.1.2 Programming Behaviour Model

Behaviour model was programmes here using the following values:

- Very recommended behaviours (++) have values of 0.5;
- Recommended behaviours (+) are associated to value 0.3;
- Unconcerned behaviours (∘) have values of 0.2;
- Unwise behaviours present values of 0.1;
- Forbidden behaviours (×) show values of 0.01.

Once behaviour tables are filled with values according to available knowledge, they are normalized before building joint distribution.

H.2 Prey Elementary Filter

H.2.1 Relevant Variables

$$S^j_{prey} = \{Y^j_{pres} Y^j_{dist} Y^j_{dir}\}$$

$$Z^j_{prey} = \{ZY^j_{pres} ZY^j_{dist} ZY^j_{dist}\}$$

Coherence based fusion variables are: λ^j_{prey}, τ^j_{prey}, β^j_{prey}

H.2.2 Joint Distribution

$$\begin{aligned}
&P(S^{0:t}_{prey} Z^{0:t}_{prey} ZD^{0:t}_{pan} M^{0:t}_{rot} M^{0:t}_{trans} M^{0:t}_{cons} B^{0:t} \lambda^{0:t}_{prey} \tau^{0:t}_{prey} \beta^{0:t}_{prey} | \pi_{prey}) \\
&= \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(Y^j_{pres} | Y^{j-1}_{pres} \pi_{prey}) \times P(Y^j_{dir} | Y^{j-1}_{dir} B^{j-1} M^{j-1}_{rot} \pi_{prey}) \\ \times P(Y^j_{dist} | Y^{j-1}_{dist} B^{j-1} \pi_{prey}) \\ \times P(ZD^j_{pan} | \pi_{prey}) \times P(ZY^j_{pres} | Y^j_{pres} Y^j_{dir} ZD^j_{pan} \pi_{prey}) \\ \times P(ZY^j_{dir} | Y^j_{dir} ZY^j_{pres} ZD^j_{pan} \pi_{prey}) \\ \times P(ZY^j_{dist} | Y^j_{dist} Y^j_{dir} ZY^j_{pres} ZD^j_{pan} \pi_{prey}) \\ \times P(B^j | \pi_{prey}) \times P(\beta^j_{prey} | B^j B^{j-1} Y^j_{pres} Y^j_{dist} \pi_{prey}) \\ \times P(M^j_{rot} | \pi_{prey}) \times P(\lambda^t_{prey} | M^j_{rot} B^j Y^j_{dir} \pi_{prey}) \\ \times P(M^j_{trans} | \pi_{prey}) \times P(\tau^t_{prey} | M^j_{trans} B^j Y^j_{dist} \pi_{prey}) \\ \times P(M^j_{cons} | M^{j-1}_{cons} B^j Y^j_{pres} Y^j_{dist} Y^j_{dir} \pi_{prey}) \end{array} \right] \\ \times P(S^0_{prey} Z^0_{prey} ZD^0_{pan} \lambda^0_{prey} \tau^0_{prey} M^0_{rot} M^0_{trans} M^0_{cons} B^0 \beta^0_{prey} | \pi_{prey}). \end{array} \right.
\end{aligned}$$

H.2.3 Parametrical Forms and Parameters Identification

Dynamic model

for presence consists in a dilution over time, as shown in table H.1.

For distance, it is considered that if robot is chasing prey ($B^{j-1} = Chase$), its distance reduces, as shown in table H.2. For all other behaviours, dynamic model is defined as:

$$P(Y^j_{dist} | Y^{j-1}_{dist} [B^{j-1} \neq CHASE] \pi_{prey}) = \mathcal{G}(Y^j_{dist}), \mu = Y^{j-1}_{dist}, \sigma = 0.5.$$

For direction, it is necessary to correct the direction in relation to robot movement in previous time step. If robot is following prey, the robot displacement in previous time step was related to prey position. So, prey direction is corrected by means of a

Y^j_{pres} \ Y^{j-1}_{pres}	0	1
0	0.6	0.3
1	0.4	0.7

Table H.1: Probability table for $P(Y^j_{pres}|Y^{j-1}_{pres}\pi_{prey})$.

Y^j_{dist} \ Y^{j-1}_{dist}	0	1	2	3	4	5
0	0.88889	0.5889	1e-1	1e-2	1e-3	1e-4
1	1e-1	0.3	0.489	1e-1	1e-2	1e-3
2	1e-2	1e-1	0.3	0.48	1e-1	1e-2
3	1e-3	1e-2	1e-1	0.3	0.489	1e-1
4	1e-4	1e-3	1e-2	1e-1	0.3	0.5889
5	1e-5	1e-4	1e-3	1e-2	1e-1	0.3

Table H.2: Probability table for $P(Y^j_{pres}|Y^{j-1}_{pres}[B^{j-1} = Chase]\pi_{prey})$.

constant K multiplying rotation speed in previous time step. This constant stablishes the relation between rotation speed and prey direction. It indicates the variation on direction variable resulting from the rotation speed aplied in previous time step. The value of this constant is related to duration of one time step and how values of rotation speed are transformed in wheels turn and consequent robot displacement, and its value is identified empirically, after executing some experiments.

$$P(Y^j_{dir}|Y^{j-1}_{dist}B^{j-1}M^{j-1}_{rot}\pi_{prey}) = \begin{cases} \mathcal{G}(Y^j_{dir}), \mu = Y^{j-1}_{dist}, \sigma = 0.5, \text{if } B^{j-1} \neq Chase; \\ \mathcal{G}(Y^j_{dir}), \mu = Y^{j-1}_{dist} + K * M^{j-1}_{rot}, \sigma = 0.5, \text{if } B^{j-1} = Chase. \end{cases}$$

Sensor model

now has another observation: direction of head. This observation is not directly related to a state and it is used in all filters: its a priori must be uniform.

Presence variable, for example, can only be observed if object is inside camera field of view. Relation between prey direction and present head position is then defined by table H.3, $|T^j_{dir} - ZD^j_{pan}| \leq 2$, that is, trainer is inside visible area. Uniform distribution is applied otherwise.

Distance variable is updated if prey was detected in image ($ZY^j_{pres} = 1$) and if inside camera field of view ($|Y^j_{dir} - ZD^j_{pan}| \leq 2$).

$$P(ZY^j_{dist}|Y^j_{dist}Y^j_{dir}ZD^j_{pan}ZY^j_{pres}\pi_{prey}) = \begin{cases} \mathcal{G}(ZY^j_{dist}), \mu = Y^j_{dist}, \sigma = 0.5, \text{if } ZY^j_{pres} = 1 \text{ and } |Y^j_{dir} - ZD^j_{pan}| \leq 2; \\ \text{Uniform, otherwise.} \end{cases}$$

ZY^j_{pres} \ Y^j_{pres}	0	1
0	0.7	0.1
1	0.3	0.9

Table H.3: Probability table for $P(ZY^j_{pres}|Y^j_{pres}Y^j_{dir}ZD^j_{pan}\pi_{prey})$, if $|Y^j_{dir} - ZD^j_{pan}| \leq 2$.

B^{j-1}	E		M		C			O		W	
Y^j_{pres}	0	1	0	1	0	1		0	1	0	1
B^j \ Y^j_{dir}	*	*	*	*	0..3	4..5	*	*	*	*	*
E	+	++	+	-	+	○	+	+	-	+	-
M	+	-	+	-	+	○	+	+	-	+	-
C	×	++	×	++	×	++	++	×	++	×	++
O	+	++	+	-	+	+	○	+	-	+	-
W	+	×	+	×	+	×	×	+	×	++	×

Table H.4: Probability table for $P(B^j|B^{j-1}Y^j_{pres}Y^j_{dist}\pi_{prey})$, behaviour selection direct model for prey filter. E stands for Escape behaviour; C denotes Chase behaviour; 0 indicates Obey behaviour; and M stands for Motionless behaviour. ++ means highly recommended, + means recommended, ○ means unconcerned, - means unwise and × stands for forbidden. An asterisk is employed when situation is the same for all values of one variable.

Observation of prey direction is then determined by both direction state and head direction observation.

$$P(ZX^j_{dir}|Y^j_{dir}ZD^j_{pan}ZY^j_{pres}\pi_{prey}) = \begin{cases} \mathcal{G}(ZX^j_{dir}), \mu = Y^j_{dir}, \sigma = 0.25, \text{if } ZY^j_{pres} = 1 \text{ and } |Y^j_{dir} - ZD^j_{pan}| \leq 2; \\ \text{Uniform, otherwise.} \end{cases}$$

Behaviour model

expresses how behaviour can be selected based on information about prey. Behaviour direct model is programmed as indicated in table H.4.

Motor model

consists in choosing rotation speed and translation speed for chasing behaviour: for all others, uniform distribution is applied.

$$P(M^j_{rot}|B^j = Y^j_{dir}\pi_{prey}) = \begin{cases} \mathcal{G}(M^j_{rot}), \mu = Y^j_{dir}, \sigma = 0.5, \text{if } B^j = Chase \\ \mathcal{G}(M^j_{rot}), \mu = 0.0, \sigma = 8, \text{if } B^j = Wander \\ \text{Uniform, otherwise.} \end{cases}$$

$$P(M^j_{trans}|B^j Y^j_{dist}\pi_{prey}) = \begin{cases} \mathcal{G}(M^j_{trans}), \mu = Y^j_{dist}, \sigma = 0.5, \text{if } B^j = Chase \\ \mathcal{G}(M^j_{trans}), \mu = S_w, \sigma = 5, \text{if } B^j = Wander \\ \text{Uniform, otherwise.} \end{cases}$$

S_w is a constant equal to half of maximum translation speed.

$$P(M^j_{cons}|M^{j-1}_{cons}B^j Y^j_{pres}Y^j_{dist}Y^j_{dir}\pi_{prey}) = \begin{cases} 1, \text{if } Y^j_{pres} = 1, Y^j_{dist} = 0 \text{ and } Y^j_{dir} = 0 \\ 0, \text{Otherwise} \end{cases}$$

H.3 Trainer Elementary Filter

H.3.1 Relevant Variables

$$S^j_{trai} = \{T^j_{pres}T^j_{dist}T^j_{dir}\}$$

$$Z^j_{trai} = \{ZT^j_{pres}ZT^j_{dist}ZT^j_{dist}\}$$

Coherence based fusion variables are: λ^j_{rot}, λ^j_{trans}, β^j_{trai}

H.3.2 Joint Distribution

$$P(S^{0:t}_{trai}Z^{0:t}_{trai}ZD^{0:t}_{pan}M^{0:t}_{rot}M^{0:t}_{trans}M^{0:t}_{cons}B^{0:t}\lambda^{0:t}_{rot}\lambda^{0:t}_{trans}\beta^{0:t}_{trai}|\pi_{trai}) = \left| \begin{array}{l} \prod_{j=1}^{t} \left[\begin{array}{l} P(T^j_{pres}|T^{j-1}_{pres}\pi_{trai}) \times P(T^j_{dir}|T^{j-1}_{dir}B^{j-1}M^{j-1}_{rot}\pi_{trai}) \\ \times P(T^j_{dist}|T^{j-1}_{dist}B^{j-1}\pi_{trai}) \\ \times P(ZD^j_{pan}|\pi_{trai}) \times P(ZT^j_{dir}|T^j_{dir}ZD^j_{pan}ZT^j_{pres}\pi_{trai}) \\ \times P(ZT^j_{pres}|T^j_{pres}T^j_{dir}ZD^j_{pan}\pi_{trai}) \\ \times P(ZT^j_{dist}|T^j_{dist}T^j_{dir}ZD^j_{pan}ZT^j_{pres}\pi_{trai}) \\ \times P(B^j|\pi_{trai}) \times P(\beta^j_{trai}|B^jB^{j-1}T^j_{pres}T^j_{dist}\pi_{trai}) \\ \times P(M^j_{rot}|\pi_{trai}) \times P(\lambda^t_{rot}|M^j_{rot}B^jT^j_{dir}\pi_{trai}) \\ \times P(M^j_{trans}|\pi_{trai}) \times P(\lambda^t_{trans}|M^j_{trans}B^jT^j_{dist}\pi_{trai}) \end{array} \right] \\ \times P(S^0_{trai}Z^0_{trai}ZD^0_{pan}\lambda^0_{rot}\lambda^0_{trans}M^0_{rot}M^0_{trans}M^0_{cons}B^0\beta^0_{trai}|\pi_{trai}). \end{array} \right|$$

T^j_{pres} \ T^{j-1}_{pres}	0	1
0	0.6	0.3
1	0.4	0.7

Table H.5: Probability table for $P(T^j_{pres}|T^{j-1}_{pres}\pi_{trai})$.

T^j_{dist} \ T^{j-1}_{dist}	0	1	2	3	4	5
0	0.88889	0.5889	1e-1	1e-2	1e-3	1e-4
1	1e-1	0.3	0.489	1e-1	1e-2	1e-3
2	1e-2	1e-1	0.3	0.48	1e-1	1e-2
3	1e-3	1e-2	1e-1	0.3	0.489	1e-1
4	1e-4	1e-3	1e-2	1e-1	0.3	0.5889
5	1e-5	1e-4	1e-3	1e-2	1e-1	0.3

Table H.6: Probability table for $P(T^j_{pres}|T^{j-1}_{pres}[B^{j-1} = Obey]\pi_{trai})$.

H.3.3 Parametrical Forms and Parameters Identification

Dynamic model for presence consists in a dilution over time:

For distance, it is considered that if robot is following trainer ($B^{j-1} = Obey$), its distance reduces, as shown in table H.6. For all other behaviours, dynamic model is defined as:

$$\begin{aligned}&P(T^j_{dist}|T^{j-1}_{dist}[B^{j-1} \neq Obey]\pi_{trai})\\&= \mathcal{G}(T^j_{dist}), \mu = T^{j-1}_{dist}, \sigma = 0.5.\end{aligned}$$

For direction, it is necessary to correct the direction in relation to robot movement in previous time step.

If robot is following trainer, the robot displacement in previous time step was related to trainer. So, trainer direction is corrected by means of a constant K multiplying rotation speed in previous time step. This constant stablishes the relation between rotation speed and trainer direction. It indicates the variation on direction variable resulting from the rotation speed aplied in previous time step. The value of this constant is related to duration of one time step and how values of rotation speed are transformed in wheels turn and consequent robot displacement, and its value is identified empirically.

$$\begin{aligned}&P(T^j_{dir}|T^{j-1}_{dist}B^{j-1}M^{j-1}_{rot}\pi_{trai})\\&= \begin{cases} \mathcal{G}(T^j_{dir}), \mu = T^{j-1}_{dist}, \sigma = 0.5, \text{if } B^{j-1} \neq Obey;\\ \mathcal{G}(T^j_{dir}), \mu = T^{j-1}_{dist} + K * M^{j-1}_{rot}, \sigma = 0.5, \text{if } B^{j-1} = Obey.\end{cases}\end{aligned}$$

Sensor model now has another observation: direction of head. This observation is not directly related to a state and it is used in all filters: its a priori must be uniform.

ZT^j_{pres} \ T^j_{pres}	0	1
0	0.7	0.1
1	0.3	0.9

Table H.7: Probability table for $P(ZT^j_{pres}|T^j_{pres}T^j_{dir}ZD^j_{pan}\pi_{trai})$, if $|T^j_{dir} - ZD^j_{pan}| \leq 2$.

Presence variable, for example, can only be observed if object is inside camera field of view. Relation between trainer direction and present head position is defined by table H.7, $|T^j_{dir} - ZD^j_{pan}| \leq 2$, that is, trainer is inside visible area. Uniform distribution is applied otherwise.

Distance variable is updated if trainer is detected in image ($ZT^j_{pres} = 1$) and if inside camera filed of view ($|T^j_{dir} - ZD^j_{pan}| \leq 2$).

$$P(ZT^j_{dist}|T^j_{dist}T^j_{dir}ZD^j_{pan}ZT^j_{pres}\pi_{trai}) = \begin{cases} \mathcal{G}(ZT^j_{dist}), \mu = T^j_{dist}, \sigma = 0.5, \text{if } ZT^j_{pres} = 1 \text{ and } |T^j_{dir} - ZD^j_{pan}| \leq 2; \\ \text{Uniform, otherwise.} \end{cases}$$

Observation of trainer direction is then determined by both direction state and head direction observation.

$$P(ZT^j_{dir}|T^j_{dir}ZD^j_{pan}ZT^j_{pres}\pi_{trai}) = \begin{cases} \mathcal{G}(ZT^j_{dir}), \mu = T^j_{dir}, \sigma = 0.25, \text{if } ZT^j_{pres} = 1 \text{ and } |T^j_{dir} - ZD^j_{pan}| \leq 2; \\ \text{Uniform, otherwise.} \end{cases}$$

Behaviour model expresses how behaviour can be selected based on information about prey. Behaviour direct model is programmed as indicated in table H.8.

Motor model consists in choosing rotation speed and translation speed for Obey behaviour: for all others, uniform distribution is applied.

$$P(M^j_{rot}|B^jT^j_{dir}\pi_{trai}) = \begin{cases} \mathcal{G}(M^j_{rot}), \mu = T^j_{dir}, \sigma = 0.5, \text{if } B^j = Obey \\ \mathcal{G}(M^j_{rot}), \mu = 0, \sigma = 8, \text{if } B^j = Wander \\ \text{Uniform, otherwise.} \end{cases}$$

$$P(M^j_{trans}|B^jT^j_{dist}\pi_{trai}) = \begin{cases} \mathcal{G}(M^j_{trans}), \mu = T^j_{dist}, \sigma = 0.5, \text{if } B^j = Obey \\ \mathcal{G}(M^j_{trans}), \mu = S_w, \sigma = 8, \text{if } B^j = Wander \\ \text{Uniform, otherwise.} \end{cases}$$

B^{j-1}	E		M		C			O			W	
T^j_{pres}	0	1	0	1	0	1		0	1		0	1
T^j_{dir} / B^j	*	*	*	*	0..3	4..5	*	*	*	*	*	*
E	+	-	+	-	+	-	-	+	-	+	+	-
M	+	-	+	-	+	-	-	+	-	+	+	-
C	+	-	+	-	+	+	++	+	+	+	+	-
O	×	++	×	++	×	++	++	×	++	++	×	++
W	+	×	+	×	+	×	×	×	×	×	+	×

Table H.8: Probability table for $P(B^j|B^{j-1}T^j_{pres}T^j_{dist}\pi_{trai})$, behaviour selection direct model for trainer filter. E stands for Escape behaviour; C denotes Chase behaviour; 0 indicates Obey behaviour; and M stands for Motionless behaviour. ++ means highly recommended, + means recommended, ∘ means unconcerned, - means unwise and × stands for forbidden. An asterisk is employed when situation is the same for all values of one variable.

H.4 Predator Elementary Filter

H.4.1 Relevant Variables

$$S^j_{pred} = \{X^j_{pres}X^j_{dist}X^j_{dir}\}$$

$$Z^j_{pred} = \{ZX^j_{pres}ZX^j_{dist}ZX^j_{dir}ZX^j_{timer}\}$$

Coherence variables: λ^j_{pred}, τ^j_{esc}, and β^j_{pred}.

H.4.2 Joint Distribution

$$\begin{aligned}
&P(S^{0:t}_{pred}ZX^{0:t}_{pres}Z^{0:t}_{pred}ZD^{0:t}_{pan}M^{0:t}_{rot}M^{0:t}_{trans}B^{0:t}\lambda^{0:t}_{pred}\tau^{0:t}_{pred}\beta^{0:t}_{pred}|\pi_{pred})\\
&= \left|\begin{array}{l}
\Pi^t_{j=1}\left[\begin{array}{l}
P(X^j_{pres}|X^{j-1}_{pres}\pi_{pred}) \times P(X^j_{dir}|X^{j-1}_{dir}B^{j-1}M^{j-1}_{rot}\pi_{pred})\\
\times P(X^j_{dist}|X^{j-1}_{dist}B^{j-1}\pi_{pred})\\
\times P(ZD^j_{pan}|\pi_{pred}) \times P(ZX^j_{dir}|X^j_{dir}ZD^j_{pan}ZX^j_{pres}\pi_{pred})\\
\times P(ZX^j_{pres}ZX^j_{timer}|X^j_{pres}X^j_{dir}ZD^j_{pan}\pi_{pred})\\
\times P(ZX^j_{dist}|X^j_{dist}X^j_{dir}ZD^j_{pan}ZX^j_{pres}\pi_{pred})\\
\times P(B^j|\pi_{pred}) \times P(\beta^j_{pred}|B^jB^{j-1}X^j_{pres}X^j_{dist}\pi_{pred})\\
\times P(M^j_{rot}|\pi_{pred}) \times P(\lambda^t_{pred}|M^j_{rot}B^jX^j_{dist}X^j_{dir}\pi_{pred})\\
\times P(M^j_{trans}|\pi_{pred}) \times P(\tau^t_{pred}|M^j_{trans}B^jX^j_{dist}\pi_{pred})
\end{array}\right]\\
\times P(S^0_{pred}Z^0_{pred}ZD^0_{pan}\lambda^0_{pred}\tau^0_{pred}M^0_{rot}M^0_{trans}B^0\beta^0_{pred}|\pi_{pred}).
\end{array}\right.
\end{aligned}$$

X^j_{pres} \ X^{j-1}_{pres}	0	1
0	0.6	0.3
1	0.4	0.7

Table H.9: Probability table for $P(X^j_{pres}|X^{j-1}_{pres}\pi_{pred})$.

X^j_{dist} \ X^{j-1}_{dist}	0	1	2	3	4	5
0	0.88889	0.3	1e-1	1e-2	1e-3	1e-4
1	1e-1	0.5889	0.3	1e-1	1e-2	1e-3
2	1e-2	1e-1	0.489	0.3	1e-1	1e-2
3	1e-3	1e-2	1e-1	0.48	0.3	1e-1
4	1e-4	1e-3	1e-2	1e-1	0.489	0.3
5	1e-5	1e-4	1e-3	1e-2	1e-1	0.5889

Table H.10: Probability table for $P(X^j_{pres}|X^{j-1}_{pres}[B^{j-1} = Escape]\pi_{pred})$.

H.4.3 Parametrical Forms and Parameters Identification

Dynamic model for presence consists in a dilution over time:

For distance, it is considered that if robot is escaping predator, its distance increases, as shown in table H.10. For all other behaviours, dynamic model is defined as:

$$P(X^j_{dist}|X^{j-1}_{dist}[B^{j-1} \neq Escape]\pi_{pred}) = \mathcal{G}(X^j_{dist}), \mu = X^{j-1}_{dist}, \sigma = 0.5.$$

For direction, it is necessary to correct the direction in relation to robot movement in previous time step.

If robot is escaping from predator, the robot displacement in previous time step was related to predator positin. So, predator direction is corrected by means of a constant K multiplying rotation speed in previous time step. This constant stablishes the relation between rotation speed and predator direction. It indicates the variation on direction variable resulting from the rotation speed aplied in previous time step. The value of this constant is related to duration of one time step and how values of rotation speed are transformed in wheels turn and consequent robot displacement, and its value is identified empirically.

$$\begin{aligned} &P(X^j_{dir}|X^{j-1}_{dist}B^{j-1}M^{j-1}_{rot}\pi_{pred}) \\ &= \begin{cases} \mathcal{G}(X^j_{dir}), \mu = X^{j-1}_{dist}, \sigma = 0.5, \text{if } B^{j-1} \neq Escape; \\ \mathcal{G}(X^j_{dir}), \mu = X^{j-1}_{dist} + K * M^{j-1}_{rot}, \sigma = 0.5, \text{if } B^{j-1} = Escape. \end{cases} \end{aligned}$$

Sensor model now has another observation: direction of head. This observation is not directly related to a state and it is used in all filters: its a priori must be uniform.

$X^j_{dir}ZD^j_{pan}$	$\lvert X^j_{dir} - ZD^j_{pan}\rvert \leq 2$		$\lvert X^j_{dir} - ZD^j_{pan}\rvert > 2$	
$ZX^j_{pres}ZX^j_{timer}$ \ X^j_{pres}	0	1	0	1
0 0	0.35	0.02	0.49	0.001
0 1	0.35	0.3	0.01	0.333
1 0	0.15	0.34	0.49	0.333
1 1	0.15	0.34	0.01	0.333

Table H.11: Probability table for $P(ZX^j_{pres}ZX^j_{timer}|X^j_{pres}X^j_{dir}ZD^j_{pan}\pi_{pred})$, if $|X^j_{dir} - ZD^j_{pan}| \leq 2$.

Presence variable, for example, can only be observed if predator is inside camera field of view. Relation between predator direction and present head position is then defined by table H.11.

If predator is visible, but not present, visual observation is more probable to indicate false, as shown in the first column. If it is present inside visible area (second column), it very unlikely that both observations are false.

New observation variable has a very important role when predator is outside visible area and present (the idea is: it is there, although it can not be seen). This is shown in the last column of table and it indicates that any true combination of both observations is probable: both indicating false is highly unlikely.

Distance variable is updated if predator is detected in image ($ZX^j_{pres} = 1$) and if inside camera filed of view ($|X^j_{dir} - ZD^j_{pan}| \leq 2$).

$$P(ZX^j_{dist}|X^j_{dist}X^j_{dir}ZD^j_{pan}ZX^j_{pres}\pi_{pred}) = \begin{cases} \mathcal{G}(ZX^j_{dist}), \mu = X^j_{dist}, \sigma = 0.5, \text{if } ZX^j_{pres} = 1 \text{ and } |X^j_{dir} - ZD^j_{pan}| \leq 2; \\ \text{Uniform, otherwise.} \end{cases}$$

Observation of predator direction is then determined by both direction state and head direction observation.

$$P(ZX^j_{dir}|X^j_{dir}ZD^j_{pan}ZX^j_{pres}\pi_{pred}) \begin{cases} \mathcal{G}(ZX^j_{dir}), \mu = X^j_{dir}, \sigma = 0.25, \text{if } ZX^j_{pres} = 1 \text{ and } |X^j_{dir} - ZD^j_{pan}| \leq 2; \\ \text{Uniform, otherwise.} \end{cases}$$

Behaviour model expresses how behaviour can be selected based on information about predator. Behaviour direct model is programmed as indicated in table H.12.

H.5 Escape Route Elementary Filter

H.5.1 Relevant Variables

$$S^j_{esc} = \{E^j_{pres}E^j_{dir}\}$$

B^{j-1}	E		M			C			O,W		
X^j_{pres}	0	1	0	1		0	1		0	1	
B^j X^j_{dist}	*	*	*	0..3	4..5	*	0..3	4..5	*	0..3	4..5
E	×	++	×	++	+	×	++	+	×	++	+
M	×	-	×	-	++	×	-	++	×	-	++
C	+	-	+	-	-	+	-	+	+	-	-
0	+	+	+	+	+	+	+	++	+	+	++
W	++	×	+	×	×	+	×	×	+	×	×

Table H.12: Probability table for $P(B^j|B^{j-1}X^j_{pres}X^j_{dist}\pi_{pred})$, behaviour selection direct model for prey filter. E stands for Escape behaviour; C denotes Chase behaviour; H indicates Homing behaviour; and M stands for Motionless behaviour. ++ means highly recommended, + means recommended, ∘ means unconcerned, - means unwise and × stands for forbidden. An asterisk is employed when situation is the same for all values of one variable.

$$Z^j_{esc} = \{ZE^j_{pres}ZE^j_{dist}\}$$

Coherence based fusion variables are: β_{esc}, λ_{esc} and τ_{esc}.

H.5.2 Joint Distribution

$$\begin{aligned}
&P(S^{0:t}_{esc}Z^{0:t}_{esc}ZD^{0:t}_{pan}M^{0:t}_{rot}M^{0:t}_{trans}M^{0:t}_{cons}B^{0:t}\lambda^{0:t}_{esc}\tau^{0:t}_{esc}\beta^{0:t}_{esc}|\pi_{esc}) \\
&= \left| \begin{array}{l} \Pi^t_{j=1} \left[\begin{array}{l} P(E^j_{pres}|E^{j-1}_{pres}\pi_{esc}) \times P(E^j_{dir}|E^{j-1}_{dir}B^{j-1}\pi_{esc}) \\ \times P(ZD^j_{pan}) \times P(ZE^j_{dir}|E^j_{dir}ZD^j_{pan}ZE^j_{pres}\pi_{esc}) \\ \times P(ZE^j_{pres}|E^j_{pres}ZE^j_{dir}ZD^j_{pan}\pi_{esc}) \\ \times P(B^j|\pi_{esc}) \times P(\beta^j_{esc}|B^jE^j_{pres}\pi_{esc}) \\ \times P(M^j_{rot}|\pi_{esc}) \times P(\lambda^t_{esc}|M^j_{rot}B^jE^j_{pres}E^j_{dir}\pi_{esc}) \\ \times P(M^j_{trans}|\pi_{esc}) \times P(\tau^t_{esc}|M^j_{trans}B^jE^j_{pres}\pi_{esc}) \end{array} \right] \\ \times P(S^0_{esc}Z^0_{esc}ZD^0_{pan}\lambda^0_{esc}\tau^0_{esc}M^0_{rot}M^0_{trans}M^0_{cons}B^0\beta^0_{esc}|\pi_{esc}). \end{array} \right.
\end{aligned}$$

H.5.3 Parametrical Forms and Parameters Identification

Dynamic model for presence consists in a dilution over time:

For distance, it is considered that if robot is escaping predator, escape route distance reduces, as shown in table H.14. For all other behaviours, dynamic model is defined as:

$$P(X^j_{dist}|X^{j-1}_{dist}[B^{j-1} \neq Escape]\pi_{esc}) = \mathcal{G}(X^j_{dist}), \mu = X^{j-1}_{dist}, \sigma = 0.5.$$

For direction, it is necessary to correct the direction in relation to robot movement in previous time step.

E^j_{pres} \ E^{j-1}_{pres}	0	1
0	0.6	0.3
1	0.4	0.7

Table H.13: Probability table for $P(E^j_{pres}|E^{j-1}_{pres}\pi_{esc})$.

E^j_{dist} \ E^{j-1}_{dist}	0	1	2	3	4	5
0	0.88889	0.5889	1e-1	1e-2	1e-3	1e-4
1	1e-1	0.3	0.489	1e-1	1e-2	1e-3
2	1e-2	1e-1	0.3	0.48	1e-1	1e-2
3	1e-3	1e-2	1e-1	0.3	0.489	1e-1
4	1e-4	1e-3	1e-2	1e-1	0.3	0.5889
5	1e-5	1e-4	1e-3	1e-2	1e-1	0.3

Table H.14: Probability table for $P(E^j_{pres}|E^{j-1}_{pres}[B^{j-1} = Escape]\pi_{esc})$.

$$
\begin{aligned}
&P(X^j_{dir}|X^{j-1}_{dist}B^{j-1}\pi_{esc}) \\
&= \begin{cases} \mathcal{G}(X^j_{dir}), \mu = X^{j-1}_{dist}, \sigma = 0.5, \text{if } B^{j-1} \neq Escape; \\ \mathcal{G}(X^j_{dir}), \mu = X^{j-1}_{dist} + K, \sigma = 0.5, \text{if } B^{j-1} = Escape. \end{cases}
\end{aligned}
$$

Sensor model now has another observation: direction of head. This observation is not directly related to a state and it is used in all filters: its a priori must be uniform.

$$
\begin{aligned}
&P(ZE^j_{dir}|E^j_{dir}ZD^j_{pan}ZE^j_{pres}\pi_{esc}) \\
&= \begin{cases} \mathcal{G}(ZE^j_{dir}), \mu = E^j_{dir}, \sigma = 0.25, \text{if } ZE^j_{pres} = 1 \text{ and } |E^j_{dir} - ZD^j_{pan}| \leq 2; \\ \text{Uniform, otherwise.} \end{cases}
\end{aligned}
$$

$$
P(ZE^j_{pres}|E^j_{pres}ZE^j_{dir}ZD^j_{pan}\pi_{esc}) =
$$

Behaviour model expresses how behaviour can be selected based on information about presence of an escape route. Behaviour direct model is programmed as indicated in table H.15.

Motor model consists in choosing rotation speed and translation speed for escaping behaviour.

$$
\begin{aligned}
&P(M^j_{rot}|B^jE^j_{pres}E^j_{dir}\pi_{esc}) \\
&= \begin{cases} \mathcal{G}(M^j_{rot}), \mu = 0, \sigma = 5, \text{if } B^j = Escape \text{ and } E^j_{pres} = 0; \\ \mathcal{G}(M^j_{rot}), \mu = E^j_{dir}, \sigma = 2, \text{if } B^j = Escape \text{ and } E^j_{pres} = 1; \\ \text{Uniform, otherwise.} \end{cases}
\end{aligned}
$$

B^j \ E^j_{pres}	0	1
E	×	+
M	∘	∘
C	∘	∘
O	∘	∘
W	∘	∘

Table H.15: Probability table for $P(B^j|E^j_{pres}\pi_{esc})$, behaviour selection direct model for escape route filter. E stands for Escape behaviour; C denotes Chase behaviour; H indicates Homing behaviour; and M stands for Motionless behaviour. ++ means highly recommended, + means recommended, ∘ means unconcerned, - means unwise and × stands for forbidden. An asterisk is employed when situation is the same for all values of one variable.

$$P(M^j_{trans}|B^j E^j_{pres}\pi_{esc}) = \begin{cases} \mathcal{G}(M^j_{trans}), \mu = S_w, \sigma = 4, \text{if } B^j = Escape \text{ and } E^j_{pres} = 0; \\ \mathcal{G}(M^j_{trans}), \mu = S_{max}, \sigma = 3, \text{if } B^j = Escape \text{ and } E^j_{pres} = 1; \\ \text{Uniform, otherwise.} \end{cases}$$

H.6 Obstacle Avoidance

It is desired to control the robot using translation (V_{trans}) and rotation speed (V_{rot}) avoiding the obstacles detected in front of the robot. We assume the robot has to follow references of commands provided by an upper level module, according to desired behaviour. The variables describing these commands are called M_{rot}, for rotation speed reference variable, and M_{trans}, for translation speed reference variable.

BIBA robot has a laser sick range sensor that provides 720 data for 180 degrees, where each data gives distance from the sensor to the nearest object detected. It is too much data to be processed, and we decided to cut the 180 degrees of sensor detection in 8 zones, 20 degrees each. Each zone contains 90 measures, but only the smallest one is used.

Figure 8.10 shows that each zone is in fact considered one object located at the smaller distance measured. d_1 is the distance of obstacle in zone 1, and if no obstacle is present, this measurement will have the bigger possible value.

Obstacle avoidance method proposed here consists in one Bayesian Program for each zone, and one Bayesian Program to combine indications of all zones. The Bayesian program for each zone determines robot commands in order to avoid the obstacle detected in this zone. Fusion program then integrates commands from all zones and reference values from upper level module.

Method employed here was developed initially for a car-like robot and it was described by Koike et al. [2003b,a]. Latter works linked obstacle avoidance to navigation [Pradalier et al., 2003c, 2004, 2005].

Zone Bayesian Program

Distance measured in zone is a sensor observation , indicating how near is the detected object in this zone. A distant object does not need to be avoided, and then no special motor command is recommended.

A near object demands deviation: depending on its position, some directions are forbidden; depending on its distance, robot must move slower or even stop completely.

Bayesian program for zone i uses three relevant variables: sensor measure D_i, rotation speed V_{rot} and translation speed V_{trans}. Rotation and translation speed variables have the same type as M_{rot} and M_{trans} used in the Bayes Filter. D_i variable values the distance in centimetres, from zero to twenty: if an object is beyond two meters, D_i is equal to twenty.

Joint distribution is:

$$P(V_{rot}\ V_{trans} D_i|\pi_i) \tag{H.1}$$

$$= P(D_i|\pi_i) \times P(V_{rot}|D_i\pi_i) \times P(V_{trans}|D_i\pi_i). \tag{H.2}$$

No a priori about where the objects can be located results in an uniform distribution for $P(D_i|\pi_i)$. Rotation speed depends on object distance because a very near object requires a bigger rotation speed. The curve for $P(V_{rot}|D_i\pi_i)$ for zone 3 is given in figure 8.11. It can be seen that rotation speeds leading the robot towards the obstacle are forbidden. Translation speed is bigger when object is far, and smaller when it is near, as shown in figure 8.11. This curve is employed for all zones.

It is important to notice that curves mentioned above do not indicates which direction follow, but rather impose constraints in relation to position of an obstacle in the zone.

Utilisation of this program consists in the following question:

$$P(D_i|V_{rot}V_{trans}\pi) \propto P(V_{rot}|D_i\pi_i) \times P(V_{trans}|D_i\pi_i). \tag{H.3}$$

This question will be used by the fusion program.

Fusion Bayesian Program

In this program, motor commands from all zones are combined with reference values from upper level modules. Relevant variables here are then: set of sensor measures of eight zones $D_1 D_2 \ldots D_8$; rotation speed V_{rot};translation speed V_{trans}; M_{rot}, rotation speed reference variable; and M_{trans}, translation speed reference variable.

All commands proposed by the zones are combined with reference commands by the joint distribution shown below:

$$
\begin{aligned}
& P(M_{rot}\ M_{trans}\ V_{rot}\ V_{trans}\ D_1 \ldots D_8|\pi) \\
& = \left| \begin{array}{l} P(M_{rot}M_{trans}|\pi) \times P(D_1 \ldots D_8|\pi) \\ \times P(V_{rot}|M_{rot}\pi) \times P(V_{trans}|M_{trans}\pi) \\ \prod_{i=1}^{8} P(D_i|V_{rot}V_{trans}\pi) \end{array} \right.
\end{aligned} \tag{H.4}
$$

There is no a priori knowledge about reference values for rotation and translation speed, neither for position and direction of obstacles, so terms $P(M_{rot}M_{trans}|\pi)$ and $P(D_1 \ldots D_8|\pi)$ are uniform distributions.

Next terms represent a fusion of reference influence and obstacles avoidance evidence from each zone. Rotation speed probability distribution term $P(V_{rot}|M_{rot}\pi)$ is a Gaussian, whose mean value is reference for rotation speed with constant standard deviation. The same principle is applied to translation speed term $P(V_{trans}|M_{trans}\pi)$.

Terms $P(D_i|V_{rot}V_{trans}\pi)$ are actually questions asked to each zone joint distribution, as in equation H.3.

Utilisation of the above joint distribution consists in asking the question:

$$P(V_{rot}\ V_{trans}|M_{rot}\ M_{trans}\ D_1 \ldots D_8\ \pi),$$

where M_{rot} and M_{trans} values are the result of Bayes Filter draw over motor commands global distribution, and $D_1 \ldots D_8$ are calculated knowing measures from Sick Laser range sensor.

H.7 Utilisation

The following sequence of questions is asked at each time step:

1. Prediction question is asked to all elementary filters;
2. Behaviour questions is asked to all elementary filters
3. Fusion of answers to behaviour question from all elementary filters;
4. Value of behaviour variable is drawn from fusion result;
5. Estimation question is asked to all elementary filters;
6. Motor question is asked to all elementary filters;
7. Fusion of answers to motor question from all elementary filters;
8. Values of motor command reference variables are drawn from fusion result;
9. Obstacle avoidance question is asked to obstacle avoidance Bayesian program;

10. Values of motor command variables are drawn from obstacle avoidance program answer;

11. Update of all elementary filters with estimation states distribution and drawn values.

Bayesian Approach to Action Selection and Attention Focusing An Application in Autonomous Robot Programming

Autonomous sensory-motor systems, situated in dynamic environments, must continuously answer the ultimate question: how to control motor commands knowing sensory inputs? Solving this question is a very complex problem, because a huge flow of information must be treated under several restrictions: real-time constraints, bounded memory space, and limited processing power. One additional and major challenge is to deal with incomplete and imprecise information, usually occurring in dynamic environments. In this thesis, we address the problem of controlling autonomous sensory-motor systems and propose a succession of cumulative hypotheses and simplifications. They are defined within a precise and strict mathematical framework, called Bayesian programming, an extension of Bayesian networks. This succession consists of five stages: Utilisation of internal states; First-order Markov assumption, stationarity and Bayesian filters; Exploiting partial independence; Addition of behaviour selection mechanism; and addition of Attention focusing. The validity of these hypotheses can be discussed and challenged. However, the corresponding inferences and resulting decisions are derived mathematically. Each description of a stage is followed by its analysis according to memory requirement, processing complexity, and difficulty of modelling. Further discussions regarding robot programming and cognitive modelling points of view are also presented. Finally, we describe an implementation on a mobile robot. The results demonstrate that the proposed framework is adequate for practical purposes.

Keywords: Autonomous Robotics, Probabilistic Model, Action Selection, Selective Perception

Approche Bayésienne pour la Sélection de l'Action et la Focalisation de l'Attention. Application à la Programmation de Robots Autonomes

Les systèmes sensorimoteurs autonomes, placés dans des environnements dynamiques, doivent répondre continuellement à la question ultime : comment contrôler les commandes motrices à partir des entrées sensorielles ? Répondre à cette question est un problème très complexe, principalement à cause de l'énorme quantité d'informations qui doit être traitée, tout en respectant plusieurs restrictions : contraintes de temps réel, espace mémoire restreint, et capacité de traitement des données limitée. Un autre défi majeur consiste à traiter l'information incomplète et imprécise, habituellement présente dans des environnements dynamiques. Cette thèse s'intéresse au problème posé par la commande des systèmes sensorimoteurs autonomes et propose un enchaînement d'hypothèses et de simplifications. Ces hypothèses et simplifications sont définies dans un cadre mathématique précis et strict appelé programmation bayésienne, une extension des réseaux bayésiens. L'enchaînement se présente en cinq paliers : utilisation d'états internes ; les hypothèses de Markov de premier ordre, de stationnarité et les filtres bayésiens ; exploitation de l'indépendance partielle entre les variables d'état ; addition d'un mécanisme de choix de comportement ;la focalisation de l'attention guidée par l'intention de comportement. La description de chaque étape est suivie de son analyse selon les exigences de mémoire, de complexité de calcul, et de difficulté de modélisation. Nous présentons également des discussions approfondies concernant d'une part la programmation des robots et d'autre part les systèmes cognitifs. Enfin, nous décrivons l'application de ce cadre de programmation à un robot mobile.

Mots Clés : Robotique Autonome, Modèle Probabiliste, Sélection de l'action, Perception Sélective