



HAL
open science

Estimating Expertise from Eye Gaze and Emotions

Thomas Guntz

► **To cite this version:**

Thomas Guntz. Estimating Expertise from Eye Gaze and Emotions. Artificial Intelligence [cs.AI]. Université Grenoble Alpes [2020-..], 2020. English. NNT: 2020GRALM030 . tel-03026375

HAL Id: tel-03026375

<https://theses.hal.science/tel-03026375>

Submitted on 26 Nov 2020

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.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : Mathématiques et Informatique

Arrêté ministériel : 25 mai 2016

Présentée par

Thomas GUNTZ

Thèse dirigée par **James L. CROWLEY**, Professeur, Université Grenoble Alpes
et codirigée par **Dominique VAUFREYDAZ**, Maître de Conférences, Université Grenoble Alpes

préparée au sein du **Laboratoire d'Informatique de Grenoble**
dans l'**École Doctorale Mathématiques, Sciences et technologies de l'information, Informatique**

Estimation du Niveau d'Expertise à partir du Regard et des Émotions

Estimating Expertise from Eye Gaze and Emotions

Thèse soutenue publiquement le **07 septembre 2020**,
devant le jury composé de :

Monsieur Laurent Besacier

Professeur, Univ. Grenoble Alpes, Président

Monsieur Mohamed Chetouani

Professeur, Sorbonne Université, Rapporteur

Monsieur Marc Cavazza

Professeur, University of Greenwich, Rapporteur

Monsieur Thomas Schack

Professeur, Bielefeld University, Examineur

Monsieur James L. Crowley

Professeur, Grenoble Institut Polytechnique, Directeur de thèse

Monsieur Dominique Vaufreydaz

Maître de conférences, Univ. Grenoble Alpes, Co-directeur de thèse



Abstract

Current paradigms for human-computer interaction view the computer as a tool: a device used to reliably, efficiently, and durably perform desired functions in a manner that conforms to standards. Reliable and predictable performance preclude adaptation to the user. However, popular culture has long presented an alternative view of computers as intelligent agents capable of interacting with humans as people. This view of intelligent computers and especially intelligent robots seems to evoke an innate desire to interact with an agent that is assumed to be both intelligent and non-judgemental. The emergence of reliable technologies for machine learning and machine perception increasingly enable alternative paradigms of human-computer interaction. The emergence of a scientific community for affective computing is an example of one such alternative, in which the artificial agent perceives and evokes affection in a human companion. Other paradigms are also possible.

In this thesis, we are concerned with enabling technologies for collaborative intelligent systems. Effective collaboration requires that both the human and the computer share an understanding of their respective roles and abilities. In particular, it requires an ability to monitor the intentions and awareness of the partner in order to determine appropriate actions and behaviors. Cognitive science has much to offer in such an effort.

In recent decades, researchers in cognitive science have developed theories and models that describe human abilities for attention, awareness, understanding, and problem-solving. In this thesis, we explore how such theories can inform informatics to enable technologies for Collaborative Artificial Intelligence. In particular, we use observations of humans with different levels of expertise engaged in solving classic chess problems to explore the effectiveness of models for visual attention, awareness, understanding, and problem-solving.

We have constructed an instrument for capturing and interpreting multimodal signals of humans engaged in solving problems using off-the-shelf commercially available components combined with in-house software. Our instrument makes it possible to record body posture, gestures, facial expressions, pupil dilation, eye-scan, and fixation, as well as player interactions with the chess problem. When combined with self-reports, these recordings make it possible to construct computer models for the awareness and understanding of the game situation during problem-solving using concepts and models from cognitive science literature.

As a first experiment, chess players were recorded while engaged in problems of increasing difficulty. These recordings were used to estimate a participant's awareness of the current situation and to predict the ability to respond effectively to threats and opportunities. Analysis of the recordings demonstrates how eye-gaze, body posture, and emotional features can be used to capture and model situation awareness. This experiment validated the use of our equipment as a general and reproducible tool for the study of participants engaged in screen-based interaction involving problem-solving and suggested improvements that were possible for future experiments. These initial experiments revealed an unexpected observation of rapid changes in emotion as players attempt to solve challenging problems. Attempts to explain this observation have led us to explore the role of emotion in reasoning during problem-solving.

In the second part of the thesis, we review the literature on emotion and propose a cognitive model that describes how emotions influence the process by which subjects select chunks (concepts) for use in interpretation of a game situation. In particular, it is well known that problem-solving is strongly constrained by limits on the number of phenomena that can be considered at a time. To overcome this limit, human experts rely on abstraction to form new concepts (chunks) from

emotionally salient phenomena. Our experiments indicate that emotion plays an important role, not only in the formation of concepts but also in the selection of concepts to use in reasoning. We hypothesize that expert players retain associations of concept with emotions in long-term memory and use these to guide the selection of concepts for reasoning. This view is in accordance with Damasio's Somatic Marker hypothesis (from 1991), which posits that emotions guide behavior, particularly when cognitive processes are overloaded.

We present initial results from a follow-on experiment designed to explore the fidelity of our model and to search for evidence of the role of emotion in solving problems. Our model suggests that an association of emotions with recognized situations guides experts in their selection of partial game configurations for use in exploring the game tree.

Keywords: awareness, mental state, situation model, multimodal observation, physiological response, emotions, body gesture, eye-gaze

Résumé

Les paradigmes actuels de l'interaction homme-ordinateur considèrent l'ordinateur comme un outil : un dispositif utilisé pour exécuter de manière fiable, efficace et durable les fonctions souhaitées dans le respect des normes. Des performances fiables et prévisibles excluent toute adaptation à l'utilisateur. Cependant, la culture populaire a longtemps présenté une autre vision de l'ordinateur, celle d'un agent intelligent capable d'interagir avec l'homme en tant que personne. Cette vision des ordinateurs intelligents, et surtout des robots intelligents, semble évoquer un désir inné d'interagir avec un agent qui est supposé être à la fois intelligent et sans jugement. L'émergence de technologies fiables pour l'apprentissage et la perception des machines permet de nouveaux paradigmes d'interaction entre l'homme et l'ordinateur. L'émergence d'une communauté scientifique pour l'informatique affective est un exemple d'une telle alternative, dans laquelle l'agent artificiel perçoit et évoque l'affection chez un compagnon humain. D'autres paradigmes sont également possibles.

Dans cette thèse, nous explorons comment de telles théories issues des sciences cognitives peuvent servir de base à l'informatique pour favoriser l'émergence des technologies d'intelligence artificielle collaborative. En particulier, nous utilisons l'observation d'humains ayant différents niveaux d'expertise engagés dans la résolution de problèmes d'échecs classiques pour explorer l'efficacité des modèles pour l'attention visuelle, la prise de conscience, la compréhension et la résolution de problèmes.

Nous avons construit un instrument pour la capture et l'interprétation de signaux multimodaux d'humains engagés dans la résolution de problèmes. Notre instrument permet d'enregistrer la posture du corps, les gestes, les expressions faciales, la dilatation de la pupille et les trajectoires oculaires, ainsi que les interactions du joueur avec le problème des échecs. Combinés aux rapports verbaux des joueurs, ces enregistrements permettent de construire des modèles informatiques pour la prise de conscience et la compréhension de la situation de jeu lors de la résolution de problèmes en utilisant des concepts et des modèles issus de la littérature des sciences cognitives.

Dans le cadre d'une première expérience, les joueurs d'échecs ont été enregistrés alors qu'ils étaient engagés dans des problèmes de difficulté croissante. Ces enregistrements ont été utilisés pour estimer la conscience qu'avait un participant de la situation actuelle et pour prédire la capacité à répondre efficacement aux menaces et aux opportunités. L'analyse des enregistrements montre comment le regard, la posture du corps et les caractéristiques émotionnelles peuvent être utilisés pour capturer et modéliser la conscience de la situation. Cette expérience a validé l'utilisation de notre équipement comme outil général et reproductible pour l'étude des participants engagés dans une interaction sur écran impliquant la résolution de problèmes et a suggéré des améliorations possibles pour de futures expériences. Ces premières expériences ont révélé une observation inattendue de changements rapides dans les émotions des joueurs qui tentent de résoudre des problèmes difficiles. Les tentatives d'explication de cette observation nous ont amenés à explorer le rôle de l'émotion dans le raisonnement lors de la résolution de problèmes.

Dans la deuxième partie de la thèse, nous passons en revue la littérature sur les émotions et proposons un modèle cognitif qui décrit comment les émotions influencent le processus par lequel les sujets sélectionnent des éléments cognitifs (concepts) à utiliser dans l'interprétation d'une situation de jeu. En particulier, il est bien connu que la résolution de problèmes est fortement contrainte par les limites du nombre de phénomènes qui peuvent être considérés à la fois. Pour surmonter cette limite, les experts humains s'appuient sur l'abstraction pour former de nouveaux concepts à partir de phénomènes émotionnellement marqués. Nos expériences indiquent que l'émotion joue un rôle important, non seulement dans la formation des concepts mais aussi dans la sélection de ceux-ci dans le raisonnement. Nous émettons l'hypothèse que les experts conservent

les associations de concepts et d'émotions dans la mémoire à long terme et les utilisent pour guider la sélection des concepts pour le raisonnement. Ce point de vue est conforme à l'hypothèse du marqueur somatique de Damasio (de 1991), qui avance que les émotions guident le comportement, en particulier lorsque les processus cognitifs sont surchargés.

Nous présentons les premiers résultats d'une expérience conçue pour explorer la fidélité de notre modèle et pour rechercher des preuves du rôle des émotions dans la résolution des problèmes. Notre modèle suggère qu'une association des émotions avec des situations reconnues guide les experts dans leur sélection de configurations de jeu partielles à utiliser pour explorer l'arbre de jeu.

Mots-clés: état mental, modèle de situation, observation multimodale, réponse physiologique, émotions, gestes du corps, regard

Acknowledgements

En 2011, pendant mes études à l'IUT Informatique d'Aix-en-Provence, j'ai rencontré des enseignants-chercheurs qui m'ont marqués par leur excellence scientifique et leur grande humanité. Je pense notamment à Patricia G. et Andreaa D., qui m'ont transmis cette flamme et cette envie d'apprendre, de ne rien lâcher, d'aller toujours plus loin. Merci Andreaa D., à toi et à famille, de m'avoir accueilli pendant ces moments de doutes, de m'avoir redonné du courage et permis de continuer à avancer. Je remercie Carl H.E., professeur rencontré lors de mon séjour à KTH et pour qui j'ai énormément de respect. Sa patience et sa bienveillance envers moi ont été déterminantes dans mon choix de continuer dans le monde de la recherche.

Cette thèse a marqué un tournant dans ma vie. Je suis arrivé dans le monde de la recherche avec l'idée de collaborer avec les plus grands chercheurs pour faire un monde meilleur, de façon désintéressée. Scientifiquement, le domaine et la problématique étudiés sont une source inépuisable de questions, toutes plus intéressantes les unes des autres. Je remercie James L. Crowley et Dominique Vaudreydaz de m'avoir permis d'entrer dans le monde de la recherche, de m'avoir donner les clés pour apprendre, comprendre, évaluer, tester, comparer, progresser. Cette expérience est unique et a été très enrichissante. Merci à mes encadrants, mais aussi à l'équipe Pervasive Interaction (anciennement Prima) et à celle de Chroma.

À côté de l'équipe, Philippe D., Gérard F. et Raffaella B. ont été d'un soutien scientifique et humain inestimable. Merci à vous. Merci Raffaella, pour ta bonne humeur, tes astuces pour survivre à la pression et ton aide inestimable. J'ai adoré le temps passé ensemble à discuter.

Merci Isabelle B. et au club de l'Échiquier Grenoblois pour votre accueil et votre implication dans cette thèse.

Merci à Élodie T. et Yann L. d'avoir été là pour m'épauler.

Merci aux enseignants de l'IUT2 de Grenoble. Les vacances effectuées auprès vous, à transmettre cette flamme, m'ont permis de souffler pendant la thèse.

Merci Lionel F. et merci à Michaël I. pour avoir toujours le mot pour rire.

Le monde de la recherche m'a permis de faire des rencontres inestimables. Merci Lina M. pour ces heures à discuter ensemble, ta joie et ta bonne humeur. Tu m'as sorti de nombreux moments très difficiles. J'ai rarement rencontré des personnes aussi bienveillantes que toi. Merci mille fois.

Maxime B., nous sommes arrivés le même jour à l'INRIA, je me souviens avoir rigolé tout bas sur ta veste de costard. Cependant, je ne saurais combien te remercier d'avoir été là. Tu étais le pilier de cette équipe pour moi. Savoir que j'allais te retrouver pour discuter et rire ensemble me donner une énergie incroyable tous les jours. Tu es quelqu'un de formidable et un ami inestimable. Merci Maxime.

Nachwa A., nous avons commencé cette thèse en même temps, dans un bureau adjacent. Nous avons affronté à travers ce brouillard épais qu'est la thèse. Tu as beaucoup changé pendant ces années de thèse, tu as avancé personnellement et c'était incroyable à voir. Merci d'avoir été là, pour ton aide, ta bonne humeur et pour ta force légendaire à faire face à chaque nouveau problème

engendré la thèse jour après jour. Ce voyage aurait été bien plus triste sans toi. Merci Nachwa.

À mon copilote, à mon partenaire de B.U., à mon duo queue de la rédaction, à toi Jean P., merci. Nous avons eu le même parcours universitaire et nous nous sommes retrouvés tous les deux, dans cette B.U. pendant plus d'un an à rédiger nos thèse. Qu'il pleuve, qu'il vente, qu'il canicule, dans les meilleurs moments, mais surtout dans les pires, nous étions là. Nous étions la constante de la B.U., la constante du Martin's café. Nous nous sommes soutenus, motivés, encouragés, critiqués, maudits, mais nous n'avons rien lâché. Je ne sais pas si j'aurais pu réussir cette rédaction sans toi. Je sais que nous nous retrouverons un jour, toujours dans cette B.U. Merci à toi. Merci à la Brasserie Martin's Cafe et à la Bibliothèque Universitaire Joseph-Fourier de nous avoir accueillis.

À mes amis proches, les vrais amis qui ont toujours été là quoi qu'il arrive. Yvan G., Alysse G., Darya B., Pierre L., vous êtes des amis irremplaçables. Vous m'avez apporté un soutien et une force que je m'estime chanceux d'avoir. Merci à vous. Merci à toi Jeune Guiens, d'être une source inépuisable de force, de rire et de chlorophylle. Loué sois-tu, Jeune Guiens.

À vous, Camille C., Élodie D., Eleonore F., Charlène N., merci pour votre soutien.

À la communauté Végane de Grenoble. Votre humanité et votre bienveillance m'ont redonné de nombreuses fois de l'espoir, mais surtout un sens dans cette société qui semble avoir perdu sa boussole depuis bien longtemps.

Merci Stella D.N. et Aude R., d'avoir su m'entendre, m'écouter, me soutenir, mais surtout merci de m'avoir montré la voie à emprunter dans mon travail d'introspection pour tenir bon.

À la musique. Atlantis Chronicles, Gravity, The Dali Thundering Concept. Sacrebleu les gars, sans vous j'aurais sombré dans une sacrée folie ! L'énergie et la motivation que me procurent vos albums ont été d'une grande aide pour cette thèse. Merci à vous. Merci Frédéric P. et Ted pour ces instants musicaux.

Enfin, si je suis ici aujourd'hui, c'est grâce au soutien inconditionnel de ma famille. Merci Elisabeth pour ton temps, ton écoute et ton soutien. Merci Pierre G., Gerard G.

Merci Élodie, ma soeur, ma grande et unique soeur. Nous nous sommes beaucoup rapprochés ces dernières années. Nous venons du même endroit, mais nos chemins sont différents. Nous nous retrouvons pourtant sur tellement de points. Si on me proposait de passer autant de temps avec toi que j'en ai passé sur cette thèse, j'accepterais sans hésiter. Merci d'être ma soeur, merci d'être là, je ne serais rien sans toi.

Papa, maman, merci d'être là, chaque jour, pour soutenir inconditionnellement vos enfants. Vous êtes des parents incroyables. Je vous aime. Merci.

Papa, maman, cette thèse vous est dédiée.

Contents

1	Introduction	1
1.1	Exploring Cognition with Chess	1
1.2	Multimodal Observation of Subjects Engaged in Solving Problems	2
1.3	Thesis Outline	4
2	What Goes On When Trying to Solve a Problem?	8
2.1	Story of Two Chess Players Solving a Chess-Task	8
2.2	Background on Human Memory	9
2.2.1	Short-Term Memory	10
2.2.2	Working Memory	11
2.2.3	Long-Term Memory	12
2.3	Associative Memory	13
2.3.1	“ <i>Neurons that Fire Together Wire Together</i> ”	13
2.3.2	Bypassing Limits to Short-Term Memory with Expertise	14
2.4	Mental Models	15
2.4.1	Situation Models	16
2.4.2	Dual-Process Theory	17
2.4.3	Body Responses to Reasoning	19
2.5	Research Questions	20
3	What Is an Emotion?	21
3.1	“ <i>What Is an Emotion?</i> ”	21
3.2	Early Theories on Emotions	22
3.2.1	James-Lange’s Theory	22
3.2.2	Cannon-Bard’s Theory	23
3.2.3	Schachter-Singer’s Theory	24
3.2.4	The Appraisal Approach	25
3.2.5	Discussion	26
3.3	The “ <i>Basic Emotions</i> ” Theory	26
3.4	The Circumplex Model	30
3.5	Distinction with Moods	32
3.6	Paradox and Bias in the Study of Emotion	32
3.7	A Neurobiological Approach to Emotion	34
3.8	Emotional Memory	35
3.9	Discussion	36

4	Physiological Manifestations of Reasoning and their Observation	37
4.1	Eye-Gaze	37
4.1.1	The Human Eye	37
4.1.2	Eye-Tracking Techniques	39
4.1.3	Eye Movement Analysis	41
4.1.4	Parafoveal Vision: The Effect of Expertise on Visual Span	45
4.2	Facial Expressions	48
4.2.1	Facial Action Coding System (FACS)	48
4.2.2	Action Units (AU) Detection Systems Pipeline	49
4.2.3	From Actions Units to the Pleasure-Arousal-Dominance Model	49
4.2.4	Cognitive Interpretations	51
4.3	Body Posture	52
4.3.1	Physiological Responses to Stress	52
4.3.2	Self-Touching	54
4.3.3	Manifestations of Dominance	55
4.4	Conclusion	55
5	Observing Human Reasoning during Problem-Solving	57
5.1	Previous work	57
5.1.1	The PERCEIVER System (1969)	57
5.1.2	The MAPP System (1973)	60
5.1.3	What Can we Learn from these Systems?	62
5.2	A Multimodal Instrument to Observe Human Reasoning	63
5.2.1	Domain of Application and Limitations	63
5.2.2	Hardware	64
5.2.3	Software	65
5.3	First Experiment Involving the Local Chess Club	72
5.3.1	Experimental Hypothesis	72
5.3.2	Participants and Data Acquisition	72
5.3.3	Designing Chess Tasks	73
5.3.4	Experimental Protocol	74
5.3.5	Metrics	75
5.3.6	Pilot Experiment: Is the Experiment Protocol Valid?	76
5.4	Results	77
5.4.1	Unimodal Analysis	78
5.4.2	Statistical Classification	82
5.4.3	Features Selection	83
5.5	General Discussions	86
5.5.1	Conclusion	86
5.5.2	Drawback	87
6	Modeling Awareness from Observation of Eye-Gaze and Emotions	89
6.1	A Formal Representation of Chess Concepts and Expertise	89
6.1.1	Frames: A Formal Representation for Concepts	89
6.1.2	A Representation for Chess Concepts	91
6.1.3	A Situation Model for Chess	94
6.2	Evidence for Awareness from Observation	96
6.2.1	Computation of the Evidence from Observations	96

6.2.2	Evidence for Chunks from Fixations	97
6.2.3	Evidence for Experience from Emotions	100
6.3	Conclusion	102
7	An Experiment to Observe the Role of Emotion in Reasoning	104
7.1	Study on Self-Reporting	104
7.1.1	Verbal Analysis in Chess	105
7.1.2	The Concurrent and Retrospective Approaches	105
7.1.3	Summary	107
7.2	A Pilot Experiment	107
7.2.1	Procedure	107
7.2.2	Results and Discussions	108
7.3	A Second Experiment	110
7.3.1	Objectives and Research Questions	110
7.3.2	Participants	110
7.3.3	Designing Chess Tasks	110
7.3.4	Ontology of Chess Concepts	111
7.3.5	The Experimental Protocol	112
7.3.6	Metrics	115
7.4	Results	116
7.4.1	An Expert Player	117
7.4.2	An Intermediate Player	120
7.4.3	A Novice Player	123
7.5	General Discussion	126
8	Conclusion and Perspectives	129
8.1	Summary	129
8.2	Contributions	132
8.3	Potential Applications	133
A	Retrieve depth from 2D keypoints and Kinect’s recording	145
B	Relevant Chunks of a Task used in Experiment 2	146
C	Publications	150

List of Figures

2.1	A typical chess puzzle	9
2.2	Working Memory Model after Baddeley and Hitch (1974)	10
2.3	Embedded-Processes Model after Cowan (1999)	11
2.4	System 1 and System 2 after Kahneman (2003)	18
3.1	James’ Theory illustrated (Source: Moors, 2009).	23
3.2	The Process of Emotion Generation (Source: Reisenzein, 1983)	25
3.3	Natural Expression of Emotion Across Species, Darwin, 1872.	27
3.4	Photographies of Fore Tribesmen’ Members	28
3.5	Example of the Facial Unit Coding System on Paul Ekman.	30
3.6	Circumplex Model after Russell, 1980.	31
3.7	Emotion and Context after Barrett, Mesquita, and Gendron, 2011	32
3.8	Examples of Duchenne and Non-Duchenne smiles.	33
3.9	A Model for Emotions and Neurotransmitters	34
4.1	The Anatomy of the Human Eye	38
4.2	Visual Angle Field.	38
4.3	Rod and Cone Density Across the Retinal Surface.	39
4.4	Brain Anatomy and Visual Pathways.	40
4.5	Retinal Maps Propagation	40
4.6	A Table-Mounted Eye-Tracking System.	42
4.7	A Head-Mounted Eye-Tracking System.	42
4.8	Purkinje Images on Pupil and Calibration.	43
4.9	Check Detection Task Example.	45
4.10	Scatter-grams of Gaze Positions in the Check-Detection Task by Skill.	46
4.11	Proportion on No-saccades Trials in Check-Detection Task.	47
4.12	Proportion on Fixation in Check-Detection Task.	47
4.13	Subsets of AUs with Fully Activated Muscles.	48
4.14	OpenFace 2.0 Facial Behavior Analysis Pipeline	49
5.1	Task Presented to an Expert Chess player and PERCEIVER.	59
5.2	Recorded Eye Movements from Chess Player and PERCEIVER	59
5.3	CEEAGE Multimodal Setup	64
5.4	FaceReader Software GUI	66
5.5	OpenFace Facial Landmarks and AUs Detection	66
5.6	Tobii Pro Studio Software	68
5.7	Tobii Studio Calibration Procedure	69
5.8	OpenPose Body Detection Visualization	70
5.9	OpenPose Overall Pipeline	70

5.10	Body Keypoints Visualization	71
5.11	Multimodal Visualization of Recorded Data.	72
5.12	Example of Designed Chess Tasks.	74
5.13	Areas of Interests Designed for A Chess Tasks.	79
5.14	Histograms of Fixations.	80
5.15	Count of Emotion Variation.	81
5.16	Count of Self-Touching.	81
5.17	Standard Deviation of the Accuracy Scores for SVMs using mRMR and Fisher.	84
5.18	Self-Touches and Emotional Results.	87
6.1	Three Examples of Chess Relations.	92
6.2	A Working Memory Model after Guntz, Crowley, Vaufreydaz, Balzarini, and Dessus, 2018b	95
6.3	Toy Example with 3 Pieces	98
6.5	Scan-Gaze Example with Three Fixations: F1, F2 and F3.	99
6.6	Fixations Example with Time of Apparition	101
6.7	Possible Moves for a Novice with Recalled Affect from Long-Term Memory.	102
6.8	Possible Moves for an Expert with Recalled Affect from Long-Term Memory.	103
7.1	Think-Aloud Protocol of Good and Expert Chess Player.	106
7.2	Example of Designed N-Check-Mate and Survival Task.	111
7.3	N-Check-Mate Task with $N=3$. Black to Play.	112
7.5	Relevant chunks of N-Check-Mate ($N = 3$) Task presented using Frames.	113
7.6	Offensive and Defensive Chunks.	117
7.7	Arousal of Expert Player.	118
7.8	Scan Path of Expert between $t=0s$ to $t=5s$	118
7.9	Awareness of Relevant Chunks for Expert Player.	119
7.10	Arousal of Intermediate Player.	121
7.11	Scan Path of Intermediate between $t=10s$ to $t=20s$	121
7.12	Awareness of Relevant Chunks for Intermediate Player.	122
7.13	Arousal of Novice Player.	124
7.14	Scan Path of Novice between $t=36s$ to $t=46s$	124
7.15	Scan Path of Novice between $t=60s$ to $t=70s$	125
7.16	Awareness of Relevant Chunks for Novice Player.	125
8.1	Sketch of a Student Aware Intelligent Training Pulpit.	134
B.3	Relevant chunks of N-Check-Mate ($N = 3$) Task presented using Frames.	149

Chapter 1

Introduction

Current paradigms for human-computer interaction view the computer as a tool. This computer-centered approach relies on efficient, reliable, and predictable performance. However, popular culture has long presented an alternative view of computers as intelligent agents capable of interacting with humans as people: a human-centered paradigm. Such a view implies that the interaction between humans and computers should be natural and similar to human social interaction. Effective human social interactions rely on the successful interpretation of a variety of nonverbal communicative cues such as facial expressions, body language, gestures, postures, among other things. In particular, social interaction requires an ability to monitor the intentions and awareness of the partner in order to determine appropriate actions and behaviors.

A system that can model the awareness and attention of a user could make human-machine interaction more convenient by providing only the required information, avoiding distraction, and making the interaction more intuitive. In this thesis, we explore how theories from Cognitive Sciences can inform informatics to enable technologies for human-centered, Collaborative Artificial Intelligence. In particular, we use observations of humans engaged in solving classic chess problems to explore the effectiveness of models for visual attention, awareness, understanding, and problem-solving.

1.1 Exploring Cognition with Chess

The emergence of reliable technologies for machine learning and machine perception increasingly enables the use of alternative paradigms for human-computer interaction. In recent decades, researchers in cognitive science have developed theories and models that describe human abilities for attention, awareness, understanding, and problem-solving.

Chess has long been used in Cognitive Science to explore attention and to develop models for problem-solving. In an extensive work involving participants of all chess backgrounds, from amateurs to masters, Adriaan De Groot investigated the cognitive requirements and the thought processes involved in playing chess (De Groot, 1978). His experiments require participants to solve a chess problem and to verbalize their thought processes. This work has served as a model, inspiring an essential body of cognitive science research.

In their study, Charness et al. showed that when engaging in competitive games, chess players display engagement and awareness of the game situation with eye-gaze and fixation (Reingold, Charness, Pomplun, and Stampe, 2001; Reingold and Charness, 2005). This suggests that the mental models used by players can be at least partially determined from eye gaze, fixation, and physiological response. The ability to detect and observe such models during gameplay can provide

a new understanding of the cognitive processes that underlay human interaction.

Humans display awareness and emotions through a variety of non-verbal channels. It is increasingly possible to record and interpret information from such channels. Thanks to progress in related research, notably recently using Deep Learning approaches (Baltrusaitis, Zadeh, Lim, and Morency, 2018; Cao, Hidalgo, Simon, Wei, and Sheikh, 2018), efficient, publicly available, the software can be used to detect and track face orientation using commonly available web cameras. Concentration can be inferred from changes in pupil size (Kahneman, 2011). Measurement of physiological signs of emotion can be done by detection of Facial Action Units (Ekman, 1977; Friesen, Ekman, et al., 1983) from both sustained and instantaneous displays (micro-expressions). Heart rate can be measured from the Blood Volume Pulse as observed from facial skin color (Poh, McDuff, and Picard, 2010). Body posture and gesture can be obtained from low-cost RGB sensors with depth information (RGB+D) (Cao et al., 2018). Awareness and attention can be inferred from eye-gaze (scan path) and fixation using eye-tracking glasses as well as remote eye-tracking devices (Sheridan and Reingold, 2017). This can be directly used to reveal cognitive processes indicative of expertise (Reingold and Charness, 2005), situation awareness in human-computer interaction (HCI) systems (Losada, 2019) or vigilance system used in cars (D’orazio, Leo, and Distante, 2004).

Our goal in this thesis is to investigate the extent to which observations of eye-gaze, posture, emotion and other physiological signals can be used to model the cognitive state of subjects and to explore the integration of multiple sensor modalities to improve the reliability of detection of human displays of awareness and emotion. In particular, we use observations of humans with different levels of expertise engaged in solving classic chess problems to explore the effectiveness of models for visual attention, awareness, understanding, and problem-solving.

1.2 Multimodal Observation of Subjects Engaged in Solving Problems

We have constructed an instrument for capturing and interpreting multimodal signals of humans engaged in solving challenging problems. Such recordings can be used to reveal human awareness of the current situation and to predict the ability to respond effectively to opportunities and threats. Our instrument (see Figure 5.3 on page 64) captures eye gaze, fixations, body postures, and facial expressions signals from subjects engaged in interactive tasks on a touch screen. We use off-the-shelf cameras, a Touch-Screen computer, a remote Eye-Tracking, and two adjustable USB-LED for lighting condition control. Recording and data processing are performed using publicly available and open-source software. A wooden structure, made with a laser cutter, is used to rigidly mount the measuring equipment in order to assure identical sensor placement and orientation for all recordings.

A first pilot experiment was designed to validate our instrument and to evaluate the effectiveness of different systems and sensors for observing eye-gaze, facial expressions, body posture, pupil dilation, and cardiac rhythm. With the aid of the president of a local chess club and our partners from the CITEC laboratory, we defined 11 end-game chess problems, similar to the daily chess puzzles that can be found in magazines or on chess websites. We ordered the problems in increasing difficulty from easy to challenging tasks. Subjects were asked to solve chess tasks within a fixed, but unspecified, time period. We recorded eye gaze, facial expressions, body postures, and physiological reactions of the players as they solved problems of increasing difficulty. In a recording session with the local chess club, we recorded 21 subjects solving the problems.

Our initial hypothesis was that rapid changes of emotion correspond to the success or failure of alternative branches during game tree exploration. We observed that such recordings could be successfully used to estimate a participant’s awareness of the current situation and to predict the ability to respond effectively to challenging situations. However, we were surprised to observe that body features, and rate of change in emotion state evolved from a neutral emotion during reactive play to a period of frequent touching and rapid changes in emotion as the problems became more and more challenging. This can be partially explained knowing of memory (in particular, Working Memory) is used while reasoning. Limits on the size of Working Memory are a fundamental property of human cognition. Miller demonstrated that humans could simultaneously retain between 5 and 9 cognitive elements in Short-Term Working Memory (Miller, 1956). Most authors present Working Memory (WM) as a collection of buffers that represent propositions representing perceived phenomena or associated concepts from Long-Term Memory (LTM). Cowan (2016) reviews the history of research on Working Memory, discussing the variety of definitions and experimental demonstrations that have been used to describe this phenomenon. For chess players, the selection of the partial game description to hold in Working Memory is critical for reasoning about chess.

In order to better understand the phenomena observed in our first experiment, we have constructed a model of the cognitive processes involved, using theories from cognitive science and classic (symbolic) artificial intelligence. This model is a very partial description that allows us to ask questions and make predictions to guide our future experiments. Our model posits that experts reason with a situation model that is strongly constrained by limits to the number of entities and relations that may be considered at a time.

Situation models (Johnson-Laird, 1989) provide a formal framework for describing human comprehension and problem-solving. In logic terms, a situation model is a state graph, in which each state (situation) is defined as a logical expression of relations (predicates) defined over entities. Entities can represent observed phenomena as well as instances of concepts, procedures, or episodes from Long-term Memory. A change in the relation between entities results in a change in situation. The use of situation models provides us a framework to model the limited structure of Working Memory. This limitation forces subjects to construct abstract concepts (chunks) to describe the gameplay, in order to explore alternative moves (Miller, 1956; Simon and Chase, 1988).

Formal representations of concepts, or chunks, stored in players’ memory can be expressed as frames (Minsky, 1974). Frames define abstract concepts that can be instantiated as entities. A frame associates the entity class with a set of properties and a set of procedures that can be performed to determine the properties.

De Groot proposed that chess reasoning consists of 4 stages: Orientation, Exploration, Investigation, and Validation (De Groot, 1978). During the orientation phase, players perceive chunks that can be used to model the situation. Beginners tend to fixate on individual pieces, while expert players can be observed to fixate on the center of chunks, but make only very rapid saccades in the direction of component pieces. Small subsets of chunks are then selected and used to explore possible moves. This raises the question of how the player can select from among the large available set of chunks. We believe that emotion plays a critical role in such selection. Expert players retain associations of situations with emotions in Long-Term Memory. The rapid changes in emotion correspond to the recognition of previously encountered situations during the exploration of the game tree. We hypothesize that recalled emotions guide the selection of situation models for reasoning. This assumption is in accordance with Damasio’s Somatic Marker hypothesis, which posits that emotions guide behavior, particularly when cognitive processes are overloaded (Damasio, 1996). Recognition of situations from experience evokes emotions that are displayed as facial expressions and body posture. We hypothesize that the subject uses the evoked emotions to select from the

many possible situations for reasoning about moves during orientation and exploration.

To validate and develop our model, we have conducted a second experiment in which participants were asked to explain their reasoning. The objectives were to determine if eye-gaze, valence, arousal, and frustration could be correlated with the four phases of reasoning proposed by De Groot, and to construct an ontology for chess concepts (chunks and relations) used by players. Our model suggests that an association of emotions with recognized situations guides experts in their selection of partial game configurations for use in exploring the game tree.

1.3 Thesis Outline

The ideas in this thesis are developed as follows:

Chapter 2: What goes on when trying to solve a problem?

Chapter 2 explores previous work on human awareness with reviews of the literature on human memory, problem-solving, and mental models. To get to the heart of the matter, two fictional chess players of different levels (one novice and one expert) are described while they are engaged in solving a chess problem. Their physiological reactions, ocular trajectories, and emotional changes are observed and compared. Significant differences between these two players are noteworthy, and the following questions naturally arise from these observations: How can one explain the somatic variations of players? Are they comparable? What can we say about their awareness of the situation and comprehend the threats and opportunities?

To answer these questions, we start with a review of human memory. Human memory is composed of three interacting entities: Short-Term Memory, Working Memory, and Long-Term Memory. Short-Term Memory stores information temporarily (up to 30 seconds without rehearsal) from the perceived external world for a latter used. Working Memory is a limited buffer, gathering information from Short-Term Memory and retrieving connected information from Long-Term Memory to compute new concepts. Lastly, Long-Term Memory is an unlimited space that stores knowledge such as past experiences, procedural knowledge, or formulas.

To understand how memory is used for reasoning in solving problems, we give a review of relevant Mental Models. Situation models provide a formal framework for describing human comprehension and problem-solving (Johnson-Laird, 1989). A situation is a spatial-temporal framework composed of a list of entities, each of these holding specific properties and is associated with functional relations.

Afterward, a review on the two way of thinking is done with Kahneman's theory and his "Dual Process" Theory (Kahneman, 2011). The first one can be seen as associative thinking, whereas the second one is true reasoning. Involuntary impressions and thoughts brought up by System 1 are automatic, almost unconscious, and are created upon habits. System 2, on his side, is the central processing unit of your brain that is used to compute harder tasks that can not be handled by System 1.

The last Mental Models reviewed gives insight on somatic behavior: the Somatic Marker Theory proposed by Damasio (1996). He identified an influential link between emotional signals and this special area of the prefrontal cortex. This link explains that somatic markers, such as variation in heart rate, facial expressions, pupil dilation, body postures, muscle contractions, etc., can be observed when a stimulus or a particular situation is encountered and is engaging a cognitive process for the subject.

Chapter 2 concludes with a summary of the reviewed mental models and how they can help us to interpret the chess play

Chapter 3: What is an Emotion?

Chapter 3 examines the question of growing importance to human-computer and human-robot interaction: What is an emotion? Starting with the work of Charles Darwin, chapter 3 opens the discussion with a review of the many existing definitions of emotions, leading to the view that emotions are complex processes that affect a variety of body and brain systems. It is interesting to see the different definitions given by researchers from different fields of research: psychologists, neurobiologists, behaviorists, etc.

Several models of emotion representations are then detailed. This non-exhaustive list allows us to view and understand the evolution of knowledge about emotions. We are particularly interested in the models proposed by Ekman and Friesen: Discrete Basic Emotions (Ekman, 1957) and the one by Russel and Mehrabian Continuous emotions in 3D Model (Pleasure, Arousal, Dominance) (Russell and Mehrabian, 1977). Despite the success of these two models in predicting the physical signs of emotions from the activation of the micro-expressions of the face, these theories are subject to a certain number of criticisms (Barrett, 2011). Other models are studied, such as Barrett's, which proposes to take into account the context in which the emotion is elicited (Barrett et al., 2011).

We examine the neurobiological view that emotions are the product of the activity of neurotransmitters such as dopamine, noradrenaline, and serotonin resulting from activity in the hippocampus and amygdala (Papez, 1937; MacLean, 1990). Emotions would be the product of the activity of neurotransmitters such as dopamine, noradrenaline, and serotonin (Lövheim, 2012). We review the importance that emotions play in the retention of information and the creation of stored memories, as a result of activity in the amygdala.

This chapter sets the stage for our investigation of the role of emotions in cognition and memory during problem-solving.

Chapter 4: Physiological Manifestations of Reasoning and their Observation

Chapter 4 reviews the external and observable manifestations of cognitive processes during problem-solving, and poses the question: to what extent we can observe the reasoning of subjects engaged in solving problems? The first section summarises the importance of the gaze in observing attention. An anatomical definition of the eye is given before exploring the functioning of the human visual system. To study how gaze is used in problem-solving, we use eye-tracking techniques: infra-red cameras that track the user's gaze in order to follow the eye's fixations. We summarise several eye-tracking techniques and focus on the use of Combined Pupil and Corneal Reflection (Duchowski, 2007). We describe the hardware and software used to track gaze and summarise the measurement of gaze using fixations, scan path, and saccades (Holmqvist, Nyström, Andersson, Dewhurst, Jarodzka, and Van de Weijer, 2011). Several studies have been carried out in the framework of the chess game by observing the gaze of chess players, several great discoveries on the functioning of the brain and on the impact of expertise have been made (De Groot, 1978; Charness, Reingold, Pomplun, and Stampe, 2001; Reingold and Charness, 2005).

Facial micro-expressions can be used to infer emotional state (Ekman, 1977; Friesen et al., 1983). In particular, we describe how to measure the emotions of subjects using off-the-shelf cameras and image processing software (Den Uyl and Van Kuilenburg, 2005; Baltrusaitis et al.,

2018). These emotions are presented in two ways, in Ekman’s model (Discrete Basics Emotions, Ekman, 1957) and in Russell’s model (3D PAD model, Russell, 1980).

The final section of chapter 4 focuses on body reactions related to cognitive processes. We first discuss stressors and their bodily manifestations (Damasio, 1996; Harrigan, 1985). We see how to define stress by three different approaches: biological, phenomenological, and behavioral (Aigrain, Spodenkiewicz, Dubuisson, Detyniecki, Cohen, and Chetouani, 2016b); and what are the body features that allow us to observe the state of stress of a subject. We define self-touches in which the subject places a hand on his own body, followed by scratching, grooming, or rubbing. We see that these movement

Chapter 5: Observing Human reasoning during problem-solving

Chapter 5 describes the recording setup and the first experiment conducted for this project. The chapter begins with a discussion of similar experiments conducted by Herbert A. Simon, who analyzed the information gathering of chess players with results modeled by two programs PERCEIVER (1969) and MAPP (1973) (Simon and Barenfeld, 1969; Simon and Gilmarin, 1973). This overview gave us valuable information on early systems that mimic human reasoning.

Based on the observable metrics detailed in chapter 4, we have constructed a multimodal recording instrument to observe subjects engaged in problem-solving. We describe the hardware and software used for this instrument and present a first experiment designed to evaluate the suitability of the instrument to observe the reasoning of subjects engaged in problem-solving. This experiment served to verify that the metrics identified in chapter 4 are relevant for the study of reasoning. The results obtained presented very interesting information, validating our hypotheses, but also raised new questions about unexpected phenomena: the rapid variations in the emotional state did not correspond to our experimental hypotheses. Nevertheless, the observed metrics presented significant information to differentiate an expert chess player from an advanced player with a 90% performance using a classical machine learning (SVM) algorithm.

Chapter 6: Modeling Awareness from Observation of Eye-Gaze and Emotions

Chapter 6 presents a Cognitive model designed to explain the rapid variations in emotion observed during our first experiment. The chapter begins with a review of the Frames schema structure proposed by Minsky (1974) as a model for concepts. Several examples of representations of chess phenomena are represented using Frames. We then propose our own Working Memory model, based on the literature (reviewed in part in Chapter 2). Our model represents the limited size of Working Memory, its interactions with other memory components (episodic memory, procedural memory, and concept representation), but also its interactions with the external world (through perception and manipulative actions).

Chapter 7: an Experiment to Observe the Role of Emotion in Reasoning

In order to develop and validate our cognitive model, a second experiment is organized. First, we explain the choice of a self-reporting protocol to gather as much information as possible on the key elements used by the players during their problem-solving. To do so, we give a review of different self-reporting methods (Ericsson, 2006; Chi, 1997).

We then describe the tasks designed specifically for this experiment and the development of an ontology listing the phenomena present in these problems. This ontology helps us to correlate the players’ explanations with identified phenomena. Twenty-three volunteers recruited from a local chess club provided self-reports of awareness after solving each of and increasingly difficult set

chess problems. Results are given for players of different level: an expert, intermediate and novice player. Our analyses remain partial, but the first results are promising, and we are certain that we are on the right track to effectively analyze the reasoning of subjects engaged in problem-solving.

Chapter 8: Conclusion and Perspectives

This last chapter concludes the work carried out in this thesis by summarizing the steps taken, then proposes the many possible applications that could benefit from the contributions of our work.

Chapter 2

What Goes On When Trying to Solve a Problem?

2.1 Story of Two Chess Players Solving a Chess-Task

Picture the following scene: two chess players face a chess puzzle, displayed in Figure 2.1. The presented problem is quite common; these can be found on specialized chess websites or in magazines. In this case, it is Black to play; they are in a tough position, outnumbered, and have two of their pieces are threatened (including the Queen, a significant piece). The two players have different profiles; one is of average level who started chess three years ago and has participated in some local competitions with average results (Elo ranking¹: 1550); our second player is considered as an expert among his teammates, practicing from his early age, it is now over 15 years that he plays at both local and international competitions, he often reaches the top 5 players of the tournament (Elo ranking: 2300). Let us see how these players react when they are faced with our chess problem.

Novice's point of view - Nemo

Nemo starts to analyze the situation. His pupils are dilating. It begins by locating the essential pieces of each color: the position of Kings, Queens, Rooks, Bishops, and Knights. Then he makes a material assessment, who has the advantage in both the number and quality of pieces? After a few seconds, Nemo concludes that White, his opponent, has an advantage. During his observation, he noticed that both his Queen and his Knight are threatened. Good news, his King is well protected, as well as his opponent's. However, the main concern for Nemo is the imminent threat to his pieces. Nemo's heart rate accelerates, he stands up straight on his chair, puts his hand on his chin, frowned, blushed, and sighed slightly. Nemo lists the possible moves to save his two pieces. He quickly realizes that it is impossible, even if he checks² the opposing King to save time. Nemo puts his hand through his hair; his pupils dilate more, his face tightens and loosens as he evaluates new possibilities. He, again, reinstalls himself in his chair. He decides to save his most important

¹ The *Elo* system is a method to calculate rating for players based on tournament performance. Ratings vary between 0 and an unbounded upper bound, however the current player with the highest Elo has 2872. https://en.wikipedia.org/wiki/Elo_rating_system (last seen 01/2020)

²A **check** is a special situation in chess that occurs when a player's King is under threat during his opponent's turn. When its King is in check, the player must get out of check. If no available move is possible to move out of a check state, the game ends in **checkmate**, and the player loses. See: [https://en.wikipedia.org/wiki/Check_\(chess\)](https://en.wikipedia.org/wiki/Check_(chess)) (last seen 01/2020)

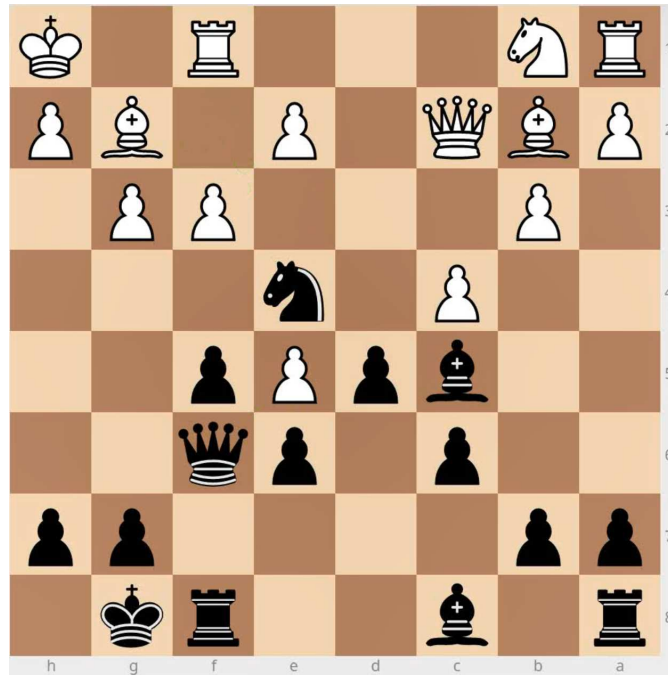


Figure 2.1: A typical chess puzzle, Black to play.

piece: his Queen. Nemo’s eyes are moving faster; he knows that he has already spent much time on the problem. Finally, He spent 5 minutes to make his decision: sacrificing his Knight to save the Queen.

Expert’s point of view - Elliott

Elliott looks at the chessboard, takes an overview of the situation, his pupils dilate slightly. “The opposing King is trapped,” he thought. Elliott is evaluating possible moves to attack the King. A slight variation in Elliott’s heart pulse is perceived, quickly followed by a minor contraction (almost imperceptible) of his facial muscles; a decision is taken. Elliott has found the solution in less than 10 seconds: check and checkmate in 3 moves.

The same problem has been presented for two players of different levels. How can one explain the somatic variations of players? Are they comparable? Can we tell, from what we see, their awareness of the situation and comprehend the threats and opportunities? To understand how players can solve a problem, we will first explore how Human Memory organizes knowledge in section 2.2. This is followed by Associative Memory and Mental Models in section 2.3 and section 2.4.

2.2 Background on Human Memory

To perform daily tasks, people must maintain access to a large amount of information. For example, a teacher might need an effort to capture the names of new students while being comfortable at giving a lesson he first learned twenty years ago. Individuals also need contextual information; readers must have access to previously mentioned characters or objects. Similarly, mental calculators must maintain intermediate results in memory. A first categorization of memory has been proposed by William James (1890) with a distinction between **primary** and **secondary** memory. The former being the small amount of information held as “*the trailing edge of the conscious*

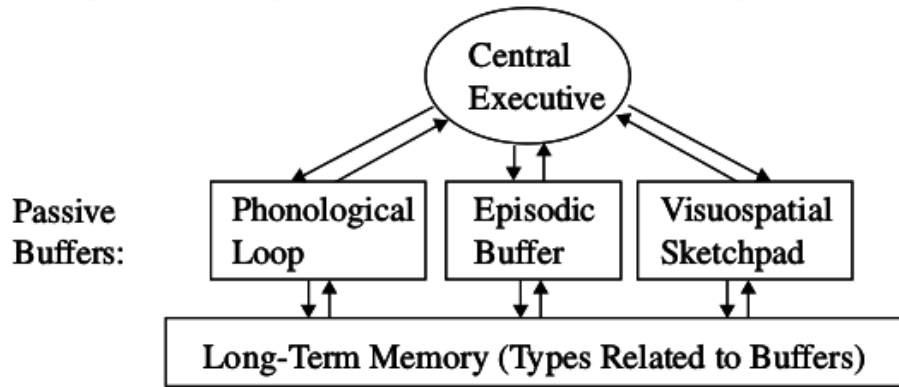


Figure 2.2: Working Memory Model after Baddeley and Hitch (1974)

present” while the latter would refer to “*the vast body of knowledge stored over a lifetime*”. In this section, we describe these two kinds of memory storage, today called Short-Term Memory (STM) and Long-Term Memory (LTM). Then, a third special storage is described later in literature: Working Memory (WM), which is used to maintain attention and perform computation.

2.2.1 Short-Term Memory

Early researchers (Broadbent, 1958; Miller, 1956; Peterson and Peterson, 1959) needed to define a special process to account for the recall of information in the short term. This special place, labeled as short-term memory, is used to temporarily retain a limited amount of information in a very accessible state. This information is not used for any manipulation, but it could be rehearsed to improve retention. For example: repeating a phone number to recall before writing it down, a small list of goods to buy at the grocery shop, etc.

Two fundamental properties that have been the subject of debate are retention duration and capacity. The duration for which information can be held in short-term memory is a challenging problem. Current theories state that information can be retained between 15 and 30 seconds if no effort is made to extend this period duration. Many cognitive processes could be used to extend retention such as the rehearsal, coding, focus of attention, association with other memory processes (Working- or Long-Term memory), etc. Covert verbal rehearsal can be used to refresh information in short-term memory. The capacity of short-term memory varies between 5 to 9 elements according to individuals’ capacity and the nature of the task (Miller, 1956; Cowan, 2016). This capacity can be extended using a special cognitive process to organize and to store information into groups, called chunks (this process is called “chunking” and will be developed in the following sections, see 2.3.2). For example, if one is familiar with acronyms of Grenoble’s research laboratories, he can easily remember the following letter series “LIGSTEEPINRIA” by decomposing it in 1- LIG (*Laboratoire d’Informatique de Grenoble*), 2- STEEP (*Soutenabilité, Territoires, Environnement, Economie et Politique* team) and 3- Inria (*Institut National de Recherche en Informatique et en Automatique*). Without meaningful association, this series of 13 letters is challenging to remember with only 4-5 units of memory (Miller, 1956; Ericsson and Kintsch, 1995).

In addition to chunking, an individual can extend the natural decay of stored information in STM by using intentional rehearsal. This is what Baddeley described in his model of memory (Baddeley and Hitch, 1974) (see Figure 2.2). He stated that a special module called the Phonological loop acts as an *inner voice* and starts to repeat any information that enters the module to cope with decaying. However, distracting tasks, or interference, can also alter retention significantly.

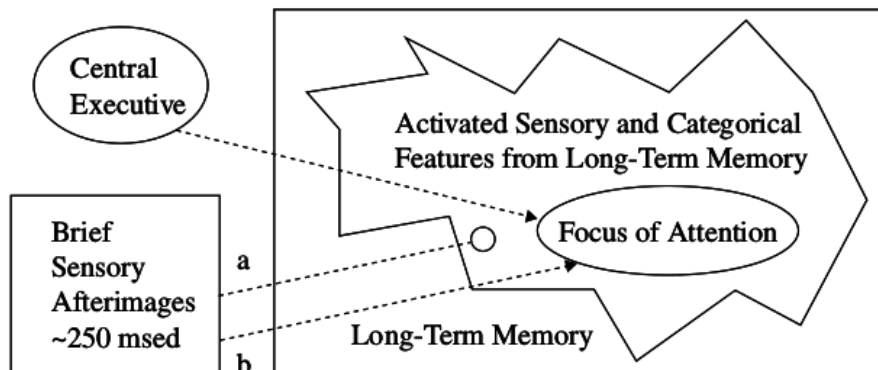


Figure 2.3: Embedded-Processes Model after Cowan (1999)

2.2.2 Working Memory

Short review on the Working Memory

Recalling a list of names and performing arithmetic are two separate mechanisms. The former only needs space to store temporary information, and can eventually be extended with rehearsal. Calculation implies storing and using partial results that can act as a distraction (Miller, Galanter, and Pribram, 1960). This suggests that a special place has to maintain the objectives to be reached and to make comparisons between possible actions; this storage was labeled as *Working Memory* and referred to the small amount of information that one can store in an accessible state for direct manipulation. Working memory can store information from stimulus as well as retrieved concepts from Long-Term Memory to be combined with the new information to obtain partial results. This concept of working memory was brought to the forefront by Baddeley and Hitch (1974); Baddeley (1992).

In his book: “*Working Memory Capacity*”, Cowan (2016) made an in-depth review of research on working memory. He concentrated on what he called the capacity of the *focus of attention*, defined as the amount of subjective or phenomenological idea of information one can retain at a given moment (see Figure 2.3).

Definition

Cowan gives a broad definition of working memory as “*the ability to remember things in an immediate-memory task (a task with no delay between the end of the presentation of items to be recalled and the period of the recall itself)*” (Cowan, 2016). Given the number of different definitions of WM, we propose to summarise the relevant concepts about WM with a single definition that we will use throughout this thesis: **WM is a collection of buffers that hold propositions representing perceived phenomena or associated concepts from LTM.**

The stored information in WM can be combined to infer new elements about the perceived world. Researchers described WM as an integrating mechanism and suggested that the limit in cognitive performance is a limit in how many dimensions of concepts can be combined (Halford, Wilson, and Phillips, 1998; Andrews and Halford, 2002). Cowan gives *Transitive Inference* as an example of an integrating mechanism: if one can observe $A > B$ and $B > C$, and if already know hierarchy concept, then he can draw the inference that $A > C$. The complexity of a concept can be measured by the number of entities that are associated. We refer to this as the arity of the concept. However, are storing concepts and processing dependant on the same resource? Halford

and colleagues stated that to avoid processing termination due to overloading of information; one could use a less complex strategy, like two-arity relations, instead of the optimal one, a three- or four-arity relation, for example. This suggests that relations between concepts can be held in WM and that more complex concepts require more energy and room to be processed.

Perceived phenomena are not the only elements stored in WM. Indeed, any new entity that enters in WM will spread energy into LTM and activate a similar concept that involves this entity (this concept is discussed later: see Section 2.3.1). Broadbent proposed that LTM always provide feedback about new information stored in WM to group them into known categories (*i.e.* chunking), these abstracts representation are stored in what is called *Semantic Memory* (Broadbent, FitzGerald, and Broadbent, 1986). Baddeley (2000) suggested that information from LTM helps to determine the form of the representation in WM. In other words, if the incoming new information is familiar, a known chunk can be used from LTM to encode it in WM. Other relevant information from LTM can also be retrieved to process new information: *Procedural Memory* provide known operations to be executed on entities to achieve a goal and *Episodic Memory* is a library of the previous recording of significant sensory memories. Communications between WM and LTM, along with perception modules, are generally modeled as a form of Hebbian active memory with propagation of activation energy (originally proposed by Hebb, 1961 and detailed in Anderson, 1983).

Distinction with Short-Term Memory

When reading theories and definitions, the line between short-term memory and working memory may appear blurry. For example, Miller et al. (1960) named a temporary and practical memory as “working memory” with no real distinction with short-term memory. Some researchers have proposed a multi-component view that describes working memory with two sub-modules: a short-term storage and a central executive processes that are used to handle stored information (Baddeley and Hitch, 1974; Cowan, 1998). In 1999, Engle et al. found unusual behaviors during tasks that required to store information before processing them while distracting stimulus are displayed (Engle, Tuholski, Laughlin, and Conway, 1999). Indeed, to improve their performance, individuals relied on their ability to inhibit non-relevant cues of the task by continually trying to look the other way from them to remained focus on their task. This correlation led Engle et al. compared working memory to “*the processes related to controlling attention*”. Conway et al. found similar results and proposed that cognitive aptitudes are directly linked with the control of attention.

In sum, in this document, we stick with the multi-components approach to describe these two distinct modules: one is a mechanism of retaining mental images of phenomenon recalled from LTM and perceived from the external world (*i.e.*, Short-Term Memory sub-module). While the second one is the central executive process that manipulates the hold information from the STM module, it is the one that drives attention and awareness (*i.e.*, Working Memory sub-module) that trigger physiological responses and visual fixations.

2.2.3 Long-Term Memory

Long-Term Memory is a vast well of wisdom where all past experiences are stored (Tulving et al., 1972). All knowledge and experiences acquired through time are assumed to be stored in this special place. This information is said to be eternal; indeed, some theorists (*e.g.*, Atkinson and Shiffrin, 1968) claimed that knowledge never disappear from memory, but instead become less and less accessible over time. The primary purpose of this memory is to retain information for later use, meaning for a retrieval period that is more than just a few seconds, which differentiate it from

the short-term memory. According to Atkinson and Shiffrin (1968) again, each time the same concept entered in short-term memory (either by rehearsal or by the repetition of the activity), it strengthens its accessibility in LTM.

Long-Term Memory is composed of different parts: the explicit (or declarative) memory and implicit (or procedural) memory. This subsection gives a definition of Long-Term Memory and its submodules.

Explicit memory

Explicit Memory, sometimes referred to as Declarative Memory, is one of the two specialized storages where factual information collected by individuals is stored. Such knowledge is, for example, past experiences, specific events, or learned concepts (Tulving et al., 1972). Recalling any of these past events are always made consciously by an individual. Explicit Memory categorizes knowledge in two ways: knowledge from past experiences and specific events (**e.g.**, last family dinner in a vegan restaurant) are stored in *Episodic Memory*; whereas general knowledge such as rules, words, formulas, algorithms, relations are organized in *Semantic Memory*. For example, the definition of Veganism (**e.g.**, the practice of abstaining from the use of animal products) is stored in Semantic Memory.

Implicit memory (or procedural memory)

Implicit memory is acquired and used unconsciously and can affect thoughts and behaviors. Its most common forms is named **Procedural Memory**, which organizes previous experiences to help people to perform specific tasks unconsciously. Implicit Memory specificity is the illusion-of-truth effect, which suggests that individuals are more likely to rate as valid familiar statements no matter how true they look.

In previous sections, we explored each category of memory: Short-Term Memory is a mechanism of retaining mental images of phenomenon recalled LTM or perceived in the external world; Working Memory is the special units that manipulate stored information in STM, drives attention as well as awareness evoking physiological responses and visual fixations; and finally, the Long-Term Memory is considered as the unlimited storage of knowledge acquired from past experiences, these knowledge are stored in different categories: explicit memory (episodic and semantic) and implicit memory. To better understand how Working Memory retrieves relevant knowledge in LTM from perceived phenomena, the next section focuses on Associative Memory.

2.3 Associative Memory

2.3.1 “Neurons that Fire Together Wire Together”

Donald O. Hebb, an influential psychologist, worked on brain surgery and human behavior at the University of McGill. He was interested in the process of learning and the potential role that neurons play in such process. After years of practice, he synthesizes his researches in his famous book “*The Organization of Behavior*” (Hebb, 1949) where he proposed a theory that described how the association of neurons are critical to the process of learning. Its breakthrough in neuropsychology led Hebb to be called the father of this area and neural networks. The theory, known as “Hebb’s rule” or cell assembly theory, is presented as followed:

Let us assume that the persistence or repetition of a reverberatory activity (or “trace”) tends to induce lasting cellular changes that add to its stability. When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A’s efficiency, as one of the cells firing B, is increased.

Hebb also stated that two cells that simultaneously get activated multiple times would tend to become *associated*. This association affects the behavior of linked neurons so that when one is activated, another one gets activated more easily the stronger the association is. This idea is the root of the artificial neural networks field. Hebb’s work directly inspires many notations and equations from today’s studies of artificial neural networks. Indeed, he described a theory to model variation of a weight shared by two neurons. Hebb’s Rule is often generalized as:

$$\Delta w_i = \rho x_i x_j \tag{2.1}$$

where w_i is the weight of a particular link between two neurons j and i , ρ is the learning rate of the weight, x_i and x_j are the input values of neuron i and j (decimal or binary value to represent activation or not as in **Hopfield** Network). The learning rate controls how strong the weight is updated using neuron activations. When the update function is linear, we can write the following:

$$y = \sum_j w_j x_j \tag{2.2}$$

From these equations, a quick interpretation that is often given by researches is “*Neurons that fire together wire together*”, this combination of neurons is seen as one processing unit because they learn to get activated together. In his book, Hebb called this combination of neurons “cell-assemblies,” and their activation is the brain’s response to stimuli. These equations are a simple way to model the human mind’s biological processes using computation machines. Even though back to the mid-1950s, computational resources were not enough to perfectly mimic human reasoning, these equations are still widely used in the artificial neural networks domain.

2.3.2 Bypassing Limits to Short-Term Memory with Expertise

In the mid-1950s, psychologists like William James agreed on the distinction between two types of memory: short- and long-term memory. However, the prevalence of behaviorism was that few (if any) experiments were focused on understanding cognitive processes, and some questions on the underlying mechanism of memory could not be tackled. George A. Miller rejected this approach in favor of a focus on speech and language. His approach aligned with Noam Chomsky, who argued that behaviorism could not explain complex human processes just by observation.

In 1956, Miller published one of the most influential papers cited in cognitive science: “The Magical Number Seven, Plus or Minus Two” (1956) where he described a curious coincidence around the number seven. While studying human cognition on different topics (immediate rehearsal of a set of digits after the presentation, recall a label while presented a distracting stimulus or fast estimation of the number of objects in a group), Miller estimated an average limit of human cognition around seven. When presented a list of items to young adults and asking them to repeat it immediately after presentation, they can correctly repeat the list if its length is between 5 to 9 items (according to individuals). He also noticed that this memory span is independent of the nature of the list item; it may be a binary digit, decimal digits or words; the magnitude is not altered. To explain this phenomenon, Miller came up with the conclusion that available slots in memory span are not of fixed length but are instead capable of encoding meaningful information. A capacity entitled “chunking”.

Chunking

Encoding information in memory using meaningful links between items is the process of “chunking”. The result of this process is called a “chunk”, a collection of basic units that share a meaningful association that can be easily stored, retrieved, and manipulates an individual’s memory. For example, when presented the following list of digits 0112358132, recalling it seems not straightforward. However, if by considering this list as a french phone number, items can be grouped two by two as: 01 12 35 81 32.

It turns out that only five groups of information are needed instead of the initial 10, which is beyond the average limit of memory span for an adult. The size of a chunk is highly personal; it relies on the nature of the information and on individuals’ ability to create a meaningful association. Meaning comes from past experiences, which implies that chunking is an ability that can be trained.

How Expertise Influences Memory

Expert telegraphers were able to lag by as much as 15-20 words when receiving Morse code (William and Harter, 1899). Chess masters can reproduce virtually an entire chess position of 32 pieces after a brief (5 seconds) presentation of a chessboard, whereas a novice can only remember the location of 3 or 4 pieces (Chase and Simon, 1973a; De Groot, 1978). How is it possible for experts to bypass the limits of short-term memory in their domain of expertise?

It appears that people create chunk when they become familiar with a set of items. A famous study conducted by Chase and Ericsson (1981) demonstrated that an adult with a standard memory span of 7 digits could be trained to increase his span significantly. After 250 hours of training (rehearsal) over a year, the trainee learned to used mnemonic systems that helped him to correctly recall up to 80 digits, which is more than ten times the normal. However, the techniques used are highly personal; they have been developed after hours of practice, and most of all, they do not generalize to other domains. This particular trainee could not remember more words than letters. A similar phenomenon has been observed for chess masters, who can not reproduce a chessboard composed of randomly placed pieces (Chase and Simon, 1973a). Indeed, without a meaningful relation between pieces, they must be stored in WM individually.

We learned from Hebb and Miller (sections 2.3.1 and 2.3.2) that associations, or cell-assemblies, are created through learning process. Connections within cell-assemblies strengthen with the repetition of activities that activate some (if not all) neurons. With times, they may become chunks: cognitive units that bond phenomenon from activities with meaningful connections. These chunks are a powerful tool that can be easily retrieved in long-term memory, manipulated in working memory, and can be intuitively perceived in real-world through pattern matching. Finally, they bypass the limited storage of short-term memory by occupying fewer slots than non-chunk elements. Given these concepts and models, we may now ask how chunks can be manipulated in memory to address everyday tasks or solve complex problems. For that, we explore in the next section different theories on Mental Models.

2.4 Mental Models

The previous sections give us an understanding of how concepts are learned and used when practicing. We discuss now how these particular cognitive processes can be represented with the framework of a situation model. To that purpose, we present in the next sections cognitive models that, we think, relevant for that particular study: we first explain Johnson-Laird’s Situation Model Theory before reviewing the distinction between two kinds of reasoning systems (System 1 and System 2) proposed by Daniel Kahneman. Finally, we examine the particular role that plays the body when reasoning through the Somatic Markers Theory, proposed by Antonio Demasio.

2.4.1 Situation Models

If the organism carries a “small-scale model” of external reality and of its possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and the future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it.

Craik (1952)

This quotation appears in the introduction to the book “*Mental Models*” by Johnson-Laird (1989). As suggested by the quotation, Johnson Laird focused on how individuals perceive, interpret and take actions within an environment. He elaborated a cognitive theory named as Situation Models that is used to represent human mental models. Over the last 25 years, theories about situation models have been adopted and developed by a large community of cognitive psychologists. As an example, Gabriel Radvansky and Jeff Zacks used Situation Models (Radvansky and Zacks, 1997) to describe readers’ mental representation to understand how information is gathered, stored and then retrieved while reading. A situation model is thus a mental representation of a described or experienced situation in a real or imaginary world.

Situation models are appropriate to represent an individual’s working memory during problem-solving. It can be used to understand, predict and evaluate possible actions from the current situation. This section defines Situation Models and describes their structure. Then, the process of information retrieving from Situation models is explained.

Description of Situations Models

Human understanding of a specific situation (real or imaginary) can be described using a situation model. The description of a Situation Models requires a formal definition of a *situation*. Radvansky and Zacks (1997), referring Perry and Barwise (1983); Barwise (1989), stated that a situation is a **spatial-temporal framework** composed of a list of **entities**, each of these holding specific **properties** and are associated with functional **relations**. More precisely, an entity can be compared to a token that is any meaningful observable phenomena such as: individual, object or idea. Each entity is described with properties that are relevant for the situation. These characteristics can refer to a physical feature (color, height, weight, etc.), to a permanent state (living, dead, in motion, etc.) or a social status (doctor, musician, woodworker, etc.). In addition, predicate functions define relations among entities and are used to structure the situation. These relations describe knowledge such as kinship, ownership, causality, spatial, temporal, etc.

Situation models are abstract representations, their instantiation depend on a person’s understanding of the analyzed situation. For example, when an individual encounters a known situation, the corresponded mental model is built upon his learned knowledge (chunks) from previous experiences stored in episodic memory. Radvansky and Zacks (1997) emphasized on the *analogue* nature of the situation model representation. This representation does not intend to be a complete and consistent mirror of the real described situation, but rather it describes the – *subjective* – meaningful structure implemented by the individual. It is important to note that this creation of mental model is driven by the goal of understanding of the individual (Graesser, Singer, and Trabasso, 1994). If no such goal exist, no mental model is created (McKoon and Ratcliff, 1992).

In his book, Johnson-Laird (1989) devoted an entire chapter on “*What is meaning?*”. We can take inspiration to define the word *meaning* as **a reference to experience** (which is in accordance with Miller’s definition, see Section 2.3.2). Experience is subjective, and so is meaning. A meaningful phenomena refers to previous encountered episodes (chunks). We also specify that experience is a gateway to prediction (**i.e.** anticipation to uncertain future event) and is used to list possible (affordable) actions. A situation model is then created when a goal of understanding a situation is defined and its structure is influenced by individual’s experiences. According to Radvansky and Zacks (1997) that situation models is stored in long-term memory, and more precisely in episodic memory. We can see it as a cell-assemblies, a hyper-chunk, regrouping a set of associated entities, properties, relations and spatial-temporal framework. Its creation and its use are both performed in working memory: the components should be reachable in a matter of milliseconds to be able to interact with the environment.

How are Situation Models created and used?

Upon construction, specific elements are integrated into the model to structure the understanding of the situation. As no model existed before, the comprehension of the scene is minimal (could even be non-existent), thus the situation contains few entities, relations and relatively small chunks. Then the model can be modified: by either adding, removing or updating information. The more the situation is understood, the more structured it became.

Retrieving information from a Situation Model stored in long-term memory require that it is brought in the foreground and become active in working memory. When facing a known situation, a memory search is activated to retrieve relevant situations (Hintzman, 1986; Radvansky and Zacks, 1997). This search is performed in according spreading activation and Hebb’s rule (Hebb, 1949; see section 2.3.1 for more information), situations that contain elements (entities, properties, relations, spatio-temporal features) from the visible situation are fired by associative activation. This allows the individual to recall all similar situation and the most meaningful is brought into working memory, as chunks. Only when a situation model is active in working memory, its structure and components became accessible. Some elements are easier to access, this is a consequence of the individual’s past experiences (interactions) with this particular situation and the way he structured it.

2.4.2 Dual-Process Theory

Kahneman argues that there are two different kinds of thinking: the first one can be seen as “*associative thinking*” whereas the second one is “*true reasoning*” (Kahneman, 2003). For example, when seeing $5+5$, one thought should directly come up to mind: 10. However, have you reasoned to get the answer? Probably not, it just popped up on your mind because there is a strong connection, built on past experiences, between these two concepts. So whenever you see the concept $5+5$, the thought of concept 10 is implicitly put forward in your mind. This is “*associative thinking*” (see section 2.3.1 on Hebb’s rule and cell-assemblies). However, when presented the calculus 14×17 , the answer is not as automatic as the first example. After reading this sentence, you most likely still do not have the answer. To do so, you would need to stop reading and engage yourself in an effortful state, thereby losing the content of your Working Memory concerning the current text. You would need to do the math. This is what Kahneman named “*true reasoning*” (by the way, the answer is 238). These two kinds of thinking, generalized as the “Dual-Process” theory, have first been proposed by William James and various researches built upon this simple idea like, among others, Wason and Evans (1974) or Sloman (1996). Kahneman proposes an interesting explanation.

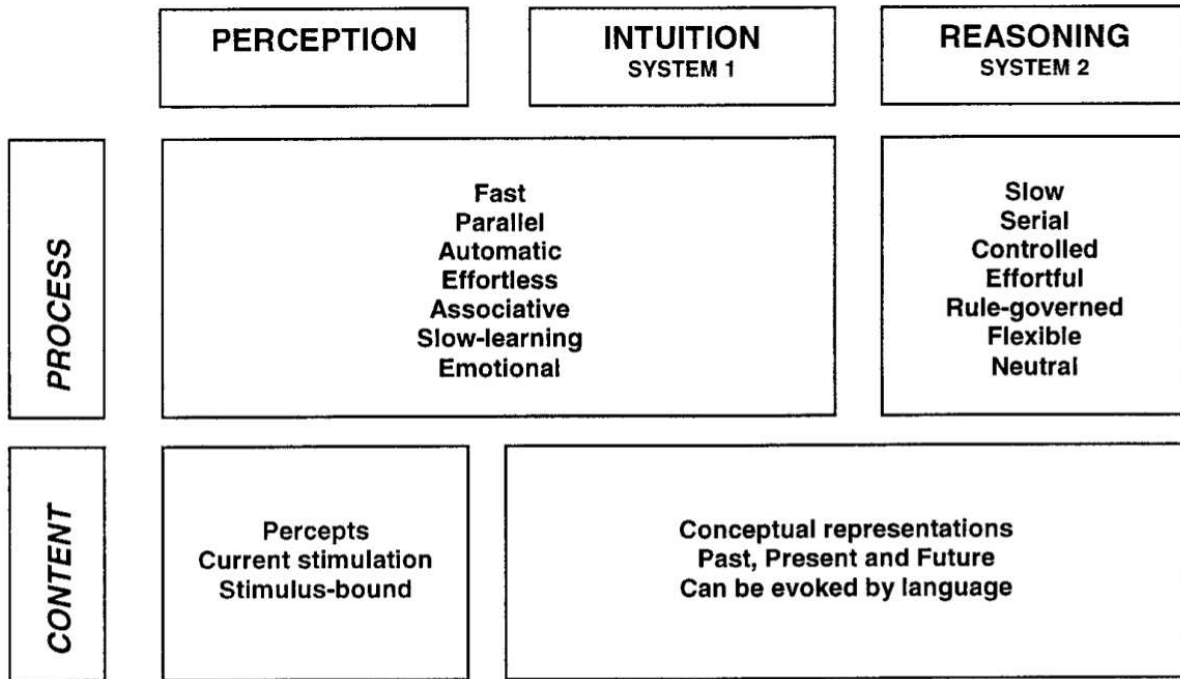


Figure 2.4: Process and Content in Two Cognitive Systems, Kahneman (2003)

System 1 and System 2

In 2003, Kahneman published a paper entitled “*A Perspective on Judgment and Choice: mapping bounded rationality*” where he reviewed his past studies done with his colleague Amos Tversky about intuitive judgment and decision making (Kahneman, 2003). He defined a detailed framework on the two different kinds of thinking that he called System 1 (associative) and System 2 (true reasoning). He and his colleague Amos Tversky were interested in the cognitive processes involved in intuitive judgments. Figure 2.4 depicts the characteristics of the two thinking systems. System 1 is our former associative thinking that is able to come up with ideas almost instantly effortless. It is described here as effortless, fast, associative, even emotional, and it presents similar features as perceptual processes. Involuntary impressions and thoughts brought up by System 1 are automatic, almost unconscious, and are created upon habits. Thus they are difficult to control. System 2, on his side, is considered as a central processing unit for the brain that is used to resolve more laborious tasks that can not be handled by System 1. It is deliberately controlled and requires the full consciousness of the individual. System 2 is also slower and effortful. Kahneman even qualified System 2 as lazy, willing to delegate as much as possible to System 1, with a tendency to accept intuitive judgments from System 1 without question directly.

A commonly used example by Kahneman to illustrate this laziness is derived from Shane Frederick’s work (personal communication, 2003), who studied cognitive self-monitoring. In this example, a simple question is asked: “*A bat and a ball cost 1.10 in total. The bat costs 1 more than the ball. How much does the ball cost?*”. Among 300 students from Princeton and the University of Michigan, less than 50% found the correct answer. Indeed, almost every student intuitively answers “10 cents” because we try to extract 1 from 1.10 (the correct answer is 0.05 cents). This easy but tricky puzzle demonstrates the lack of quality control System 2 exerts on System 1’s output.

Does this example suggest that one should not trust intuitive judgment? As we will observe several times throughout this thesis, chess masters exhibit incredible intuitive judgments in com-

petitive games (see Elliott’s performance in Section 2.1). High skill players can deliver a correct and accurate answer to the current situation in a matter of seconds (Chase and Simon, 1973b). In Kahneman (2003), the authors associate the term *intuition* with *accessibility* of knowledge. Any knowledge stored in memory that is associated with any perceived phenomena (or thought) will become more accessible (this is in accordance with Hebb, 1961, see Section 2.3.1).

2.4.3 Body Responses to Reasoning

The Somatic Marker Hypothesis

When discussing decision-making in psychology, one essential hypothesis from the neurobiologist Antonio Damasio should be mentioned: “*the Somatic Marker Hypothesis*”. Damasio was interested in understanding how the neural systems are involved in cognitive processes such as decision-making, language, memory, emotion, or even consciousness. By examining individuals with pre-frontal cortex lesions (more precisely the ventromedial pre-frontal cortex), Damasio identified an influential link between emotional signals and this special area of the pre-frontal cortex. The purpose of these signals is to regulate decision-making in complex situations (often acting non-consciously) (Damasio, 1991, 1996).

The Somatic Marker Hypothesis can be formulated following these three main assumptions:

- Both conscious and unconscious neural operations are involved in human reasoning and decision making;
- These neural operations depend on attention, working memory, and emotion;
- Reasoning and decision making depend on intuitive (**i.e.** available) knowledge stored in long-term memory that is related to the current analyzed situations, actors, actions, and outcomes.

Somatic markers are physiological signs such as variation in heart rate, facial expressions, pupil dilation, body postures, muscle contractions, etc. These body changes are responses to a stimulus or a particular encountered situation. According to Damasio, over time, these signals become a specific signature associated with a particular situation and their outcomes.

Relationship between Emotion, Memory and Decision Making

Where conventional theories on decision making addressed emotion as a consequence of a decision, Damasio proposed that physiological signals arise directly during the decision-making process (Bechara, Damasio, and Damasio, 2000). A special place of the prefrontal cortex, called Ventromedial Prefrontal Cortex (vmPFC), is associated with processes of decision making and self-control, but also with the regulation of the amygdala activity and emotions. Damasio described this brain area as a storage for learned connections between factual knowledge and bioregulatory states. Indeed, as his hypothesis suggests it, when facing a situation the vmPFC gathers (i) facts about the perceived stimuli, (ii) knowledge (factual and emotional) from related past experiences; this is made possible by the special internal structure of the vmPFC that provides groundwork for learning and retrieving association between (i) and (ii). This learning process makes sense if we analyze it following the Hebb’s rule: *Neurons that fire together wire together* (see section 2.3.1). Last but not least, these connections can be interpreted as templates, they do not store factual information but rather “*hold the potential to reactivate an emotion by acting on the appropriate*

cortical and subcortical structures” (Damasio, 1989; Bechara, Damasio, Damasio, and Anderson, 1994).

Associations of concepts seem to be recurrent in the explored Mental Models. Johnson-Laird defines his situations with entities that are bonded with meaningful relations. Kahneman describes System 1 as associative to explain the intuitive and rapid thought that comes up to mind. Associations also involve somatic responses, as proposed by Damasio. Learned concept (chunks) are associated with the affect that was evoked during the learning process. Would it be possible to induce what chunk is learned by someone just by observing his somatic responses?

2.5 Research Questions

Returning to our story with Nemo and Elliott solving a chess problem (section 2.1), we observed two different behaviors. We came up with several questions: How can one explain the somatic variations of players? Are they comparable? Can we tell, from what we see, their awareness of the situation, and their comprehension of threats and opportunities? To answer these questions, we explored how human memory is structured to store and manipulate concepts: Short-Term Memory, Working memory, and Long-Term memory. After that, we reviewed mental models and how associative memory is used to retrieve relevant knowledge according to the current situation.

How can we apply this knowledge to interpret the chess players’ behavior? We can easily claim that our expert Elliott gathered more knowledge in his long-term memory with his long experience. Was he able to identify a familiar pattern in the puzzle that helped him to find the solution? Previous sections suggest that to activate the relevant knowledge from Long-Term Memory, associated concepts must be present in working memory. Elliott has been able to solve the puzzle in less than 10 seconds. This implies that he was able to select the relevant concepts from the chessboard, put them in his working memory to activate the appropriate knowledge in long-term memory that helped him to solve the problem. He was right on the target in his first try, 10 seconds is a short time to consider many concepts. How was he able to select the right set of concepts among the 30 present pieces on the board?

Nemo is a beginner, but he has been considering possible concepts for more than 4 minutes. Unlike Elliott, he has failed in selecting the right concept. What was the clue that led Elliott to the right combination in 10 seconds? Why did Nemo fail to see this particular clue in 4 minutes? Observing players’ body reactions while solving the problem was also intriguing. Yet, could we have predicted the result of their reasoning from their somatic signs? The Somatic Marker Theory suggests that we can get insights from these body reactions, yet, we need a way to perceive these external signs.

From these observations and reviews from literature, we established a set of research questions that guided the thesis:

- What guides reasoning?
- Can we observe and model reasoning (working memory) from external signs?
- Can we build an instrument that observes external signs (and thus, reasoning)?

Damasio suggests that emotion may influence and guide reasoning. The following section will explore emotion models literature in Chapter 3. This is followed by chapter 4 on visual attention and physiological signs that can be observed. That way, following the somatic marker theory, we can capture external signs of subjects engaged in problem-solving and relate these to their reasoning.

Chapter 3

What Is an Emotion?

No doubt as long as man and all other animals are viewed as independent creations, an effectual stop is put to our natural desire to investigate as far as possible the causes of Expression.

“The Expression of the Emotions in Man and Animals.”
Charles Darwin

This chapter reviews the theories and approaches associated with the study of emotions. It begins with a discussion on how to define an emotion. Over the years, many researchers have proposed different approaches to the definition of emotions. This chapter reviews some of the more important theories in Section 3.1. A historical background of influential theories is depicted in Section 3.2. Historically influential theories such as James-Lange’s theory (1884) to Schachter’s theories (1962) are reviewed, and the role of the body and the mind in emotion generation are described. The chapter then focuses on two representation approaches of emotions: the basic emotions approach inspired by Charles Darwin’s work (Section 3.3) and a 3-dimensional model (Pleasure - Arousal - Dominance) (Section 3.4). While these two approaches have dominated the field for decades, they have also been the target of several criticisms. Major controversial topics, raised by Russel and Barrett, are discussed in Section 3.6.

Alongside these theoretical and empirical approaches of categorization, emotion can also be studied from a biological point of view: by tracing the level of hormones and activation of neurotransmitters. Interestingly, a similar classification of emotion seen in the previous Section also emerges from a neurobiological approach, as described in Section 3.7. The last Section (3.8) is a direct link between the neurobiological approach and other mechanisms in the body; we discuss then that emotions can influence the memory system (detailed in the previous Section 2.2). This Section is finally concluded with a discussion on the presented theories and approaches. We present our arguments for our choice in selecting one particular approach, among others, knowing the diversity in theories and approached.

3.1 “*What Is an Emotion?*”

Defining the concept of “*Emotion*” has always been a challenge for researchers. This recent and fashionable concept, generalised two centuries ago, is used to sometimes represents *passions*, *feelings*, *affects* or a combination of any of these terms. While emotion is used in everyday life, few

people will agree on a definition. James William dared to ask the question “*What is Emotion?*” in his famous publication in 1884, but only succeeded in starting an endless debate among the research community. Enumerating definitions about “*Emotion*” would be a long process, reformulating our own definitions would only add one to this endless stack of information. Thus, we would instead give a selection of definitions that *we think* relevant, and that may help the readers to understand better the notions used in the following sections of this document:

1. “*An amoral way as an autonomous physical or mental state characterised by vivid feeling and physical agitation*” – *From Passions to Emotions*, Dixon (2003)
2. “*A complex mix of internal signaling mechanisms that involve both the body and mind, helping us function in a reasonable, balanced, and healthy way. Emotion includes not only highly visible signals such as facial expressions, but also unseen internal signals that guide thoughts, motivations, planning, decisions, learning, and memory retrieval. Emotion works powerful influences behind the curtain of human performance.*” – *Affective Computing*, Picard (2000)
3. “*A strong feeling deriving from one’s circumstances, mood, or relationships with others*” – Oxford Dictionary¹

From these definitions, one can notice that several common concepts: the body and the mind are frequently listed, and mental state, as well as physical feelings, are often cited. It was quickly understood that “emotion” is a complex phenomenon, and it is not generated by only one system inside the body (e.g. a part of the brain or an endocrine gland).

3.2 Early Theories on Emotions

3.2.1 James-Lange’s Theory

One of the earliest theories on emotion comes from William James (1842-1910), who was an American philosopher and psychologist labeled as the “Father of American psychology” for his influential works. He is one of the two psychologists who formulated the famous emotion theory named James-Lange in the 1880s. The second name, Lange, comes from Carl Lange, a Danish physician who thought about the same emotion theory as James, but rather independently. James’ theory describes emotion in terms of a sequence of events: occurrence of an external stimulus, a somatic response, and a conscious emotional feeling.

Communities started to become interested in James’ theory in 1884 when he published an article titled: “*What Is an Emotion?*” (James, 1884). To answer his question, James started to ask another one: “*do we run from a bear because we are afraid or are we afraid because we run?*”. According to James, the first proposition is the expected answer; however, he argues:

“The hypothesis here to be defended says that this order of sequence is incorrect, that the one mental state is not immediately induced by the other, that the bodily manifestations must first be interposed between”.

Following his idea, the second proposition should be a more rational statement: “are we afraid because we run?”. To illustrate his theory, Figure 3.1 describes this particular sequence of events. In this image, one can see that the bodily response precedes the emotional experience, which is the main proposition of William James. Furthermore, the expressed emotion is based on the

¹<https://en.oxforddictionaries.com/definition/emotion> (last seen 01/2020)

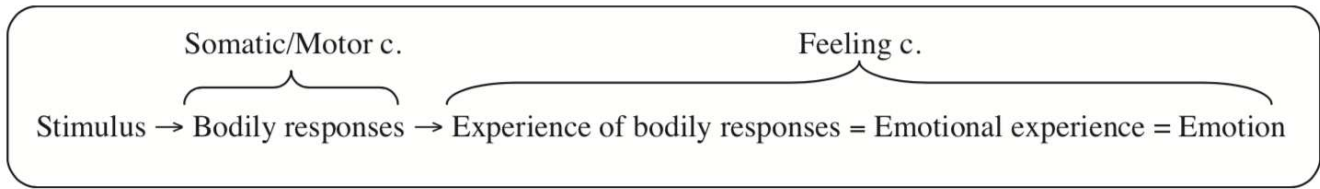


Figure 3.1: James' Theory illustrated (Source: Moors, 2009).

bodily intensity and quality responses, which means that an intensive bodily response causes a high emotion expression and vice versa.

Another interesting point raised by James is that each specific emotion has its own "response signature". *Fear* is not bodily expressed the same way as *Sad*; this explains why people can infer an individual's emotion from observation of his body responses. However, the important question in emotion theory is the elicitation problem, and James did not address it in his proposal. His theory can not answer the following question: What are the mechanisms involved to generate somatic expressions when a stimulus is displayed?

James-Lange's theory prevailed in the field for decades but started to be criticized on both on theoretical and empirical sides from the 1920s. First, Walter Cannon and his experiments (see next paragraph), then Scherer, who provided an answer to the emotion elicitation problem by decomposing it in five organismic subsystems that interact with each other. Scherer concluded that James's failure is that he could not make the distinction between "*What is an emotion?*" when he meant "*What is a feeling?*" (Scherer, 2000).

3.2.2 Cannon-Bard's Theory

In the 1920s, the psychologist Walter Bradford Cannon of Harvard University proposed with his doctoral student Philip Bard a new model for emotion in the response of James-Lange's theory. In the early 1900s, the main interest of Cannon was the study of brain's role in generating physiological responses. To do so, he and his students performed a series of pioneering studies on animal physiology and their responses to stress that led to several valuable empirical findings. For example, he is the one who coined the terms "*Fight-or-Flight*" and "*Homeostasis*".

Cannon, and Bard argued that James-Lange theory would be invalid if subjects with no visceral reaction² (from nerves' activity) display emotional expressions. Such displays would be impossible, according to James' proposal, as bodily responses and feedback are essential to emotional expressions. To this end, Cannon and his colleagues elaborated experiments in which the visceral organs and the central nervous system of cats are cut off (Cannon, 1927). In contrast to what James-Lange would predict, cats remained emotionally responsive. In addition to this finding, Cannon added that emotions are not as specific as James proposed. Indeed, with other experiments, they observed that both subjects exhibiting *fear* or *anger* come with an increase in their heart rates, which is in direct contradiction with the James's hypothesis that emotions have specific somatic responses. The last point evoked by Cannon is that artificial induction of physical arousal (e.g. by injection of adrenalin) is not interpreted as emotion (artificial or real) by the subject. Cannon and colleagues and his student Bard compiled their experimental results into a manuscript: "*The James-Lange theory of emotions: A critical examination and an alternative theory*" where they proposed a model, named as Cannon-Bard, as an alternative to the incomplete James-Lange theory (Cannon, 1927).

²Viscera are related to soft internal organs of the body

3.2.3 Schachter-Singer's Theory

The last (early) theory presented in this document is Schachter's Two-Factor Theory (Schachter and Singer, 1962). This description is inspired by the paper: *The Schachter Theory of Emotion: Two Decades Later* written by Reisenzein (1983).

Schachter was an American social psychologist who wrote several books that are still highly influential (like *Social Pressures in Informal Groups*, 1950; or *When Prophecy Fails*, 1956) and gained multiple awards for his essential findings (American Association for the Advancement of Science Socio-Psychological Prize or Prize for Behavioral Science Research among others). He obtained a position in the Columbia University faculty in New-York as a professor of psychology in 1961. Along with his student Jerome Singer, Schachter proposed a new theory based on two factors: a cognitive component and a physiological-arousal component. These two factors are used to explain in detail the different components of an emotional state and processes of its generation.

Components of an emotional state

As James theorized, Schachter and Singer define an emotional state as the product of an interaction between the body and the brain; more precisely, between physiological arousal and a cognition process focusing on the arousing situation. Each component plays an essential role in the emotional state: the physiological arousal determines its intensity while the cognition tries to determine its quality (e.g. which emotion is felt). Schachter stated that these two components are “*related multiplicatively*”: both of them are essential for the emotional state occurrence.

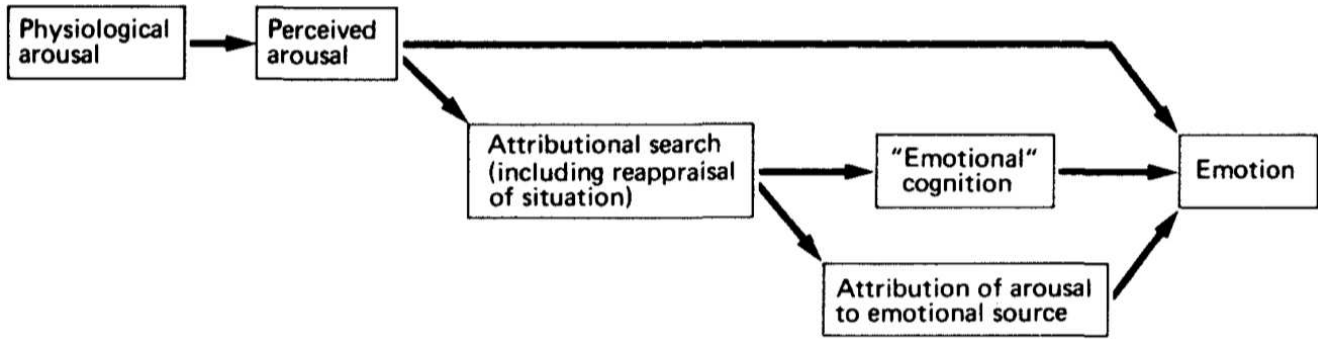
Processes in emotion generation

The emotion generation process can be broken down into two different situations: **Case 1** is the everyday life situation, where the subject knows what stimulus triggered his emotional state while **Case 2** is an atypical situation with unknown stimuli.

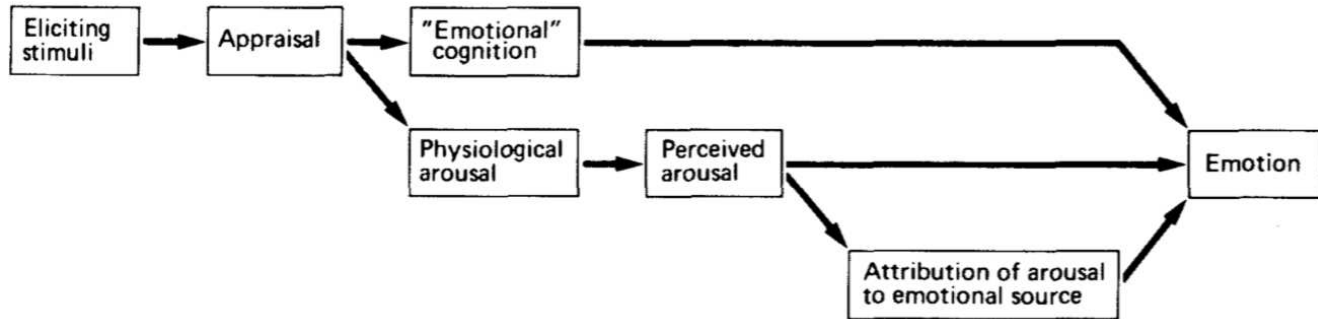
Case 1 – Schachter states that the cues provided by the known stimulus are used by the cognition component to determine the emotional state quality (see Figure 3.2a). These cognitive processes are considered unconscious for the subject, that is only aware of the sensation generated by the emotional state.

Case 2 – In this – atypical – situation, there is no immediate causal explanation for the aroused emotional state felt by the subject (see Figure 3.2b). This could happen in some particular situations (e.g. unknown side effects after drug consumption, physiological disorders, etc.), but the most important is that it can be artificially created during experiments. Schachter and Singer (1962) designed situations to observe subjects in such situations by injecting subjects with *epinephrine*, a drug known to cause an increase in blood pressure, heart, and respiration rate (Schachter and Singer, 1962). During this experiment, four conditions were used: drug-effects informed, drug-effects ignorant, drug-effects misinformed, and a control group. For those who did not receive any explanation about their sudden physiological arousal, Schachter proposed that this situation would lead to a cognitive process to search for a plausible cause. If found, a corresponding emotional state will arise for the subject. The emotion quality would only depend on the result of this attributional search process.

Initial Schachter and Singer hypothesized that both opposed situations like encountering a threatening person or a close relative would entail comparable physical arousal. Once the body responses have been generated, the sensation of the offensive emotion is attributed by the cognitive component to a cause (such as *fear* for the threat and *joy* for the reunion). Schachter and Singer supported this hypothesis with their experiment. Subjects were placed with either a happy or



(a) Case 1 – The process of emotion generation in everyday life.



(b) Case 2 – The process of emotion generation in the case of unexplained arousal.

Figure 3.2: The two cases of the process of emotion generation as described in: “*The Schachter Theory of Emotion: Two Decades Later*” (Reisenzein, 1983).

angry bystander after drug injection. It was assumed that the attributional search process of participants would interpret their arousal with the emotional state of the bystander.

This cognitive attributional process has been well-received by the scientific community. However, starting from the 1980, the Schachter-Singer’s theory has been challenged, empirically (e.g. reducing drug injection will not reduce the emotion intensity perceived by subjects, see Erdmann and Van Lindern, 1980) and theoretically (e.g. repeated exposure to stimuli led to the same emotion quality, even with a placebo injection, see Kunst-Wilson and Zajonc, 1980).

3.2.4 The Appraisal Approach

Even if Schachter-Singer’s theory has been challenged, the idea that cognition is part of the emotion generation process is appealing. A branch of emotion study has focused on this assumption and led to the Appraisal theory approach, which considers that emotion elicitation is a consequence of a cognitive evaluation (or appraisal) of the external stimulus (or a remembered event). In his book *The Emotions* (1986), the Dutch psychologist Nico Henri Frijda described the approach saying that “*emotional experience ... is experience of the situation*”. This definition entails two noteworthy facts: first, knowing the appraisal of a particular situation is enough to predict the experience emotion; secondly, the opposite is exact as well, appraisal of a situation can be deduced from the displayed emotion (Scherer, Schorr, and Johnstone, 2001). It also should be noted that the stimuli appraisal is performed unconsciously before emotion generation (and thus, before any somatic reaction). The emotion becomes conscious after its elicitation.

One interesting aspect of the Appraisal approach is the explanation of the variability of emotion elicitation in person. Indeed, theorists have used variables to explain this variation, two important ones are **goal relevance** and **goal congruence**. The former tests if the external stimulus is relevant for a subject’s goal or concern; if so, emotion is elicited. The latter variable is responsible for the nature of the emotion (positive or negative) according to the match (or mismatch) of the goal relevance and the specific stimulus. These two variables are person-centered and explain the variation of emotion elicitation across people and cultures.

3.2.5 Discussion

At the moment of writing this thesis, the scientific community does not agree on whether bodily feedback is sufficient or even necessary in the processes of emotion generation. The same conclusion applies to the cognitive component. This lack of consensus prevents us from choosing a particular theory as a basis for our model. However, what we have learned through this theory review is that emotion can be understood as multidimensional feedback observation (e.g. subjective emotional experience, somatic reactions, physiological arousal, etc.) from an internal state.

What are emotions and what are the mechanisms involved in their elicitation are, indeed, intriguing questions, and researchers are still today confronting theories. Another interesting side of emotions is their identification and classification. We will now see different approaches related to these subjects.

3.3 The “*Basic Emotions*” Theory

In his book, “*The Expression of the Emotions in Man and Animals*”, Darwin (1872) proposed that emotions and natural expressions are universal across human cultures and shared with many animal species. Darwin identified six emotional states shared among humans and animals: happiness, sadness, fear, anger, surprise, and disgust. Some of these emotions are presented in Figure 3.3. Darwin’s view has been widely supported (e.g. Allport, 1924; Tomkins, 1962; Asch, 1987) but also criticised (e.g. Klineberg, 1940; Labarre, 1947; Birdwhistell, 1970). The most extensive work – and the one that interests us here – has been performed by Paul Ekman and Wallace Friesen.

Ekman initially investigated nonverbal communication with his first publication: “*A Methodological Discussion of Nonverbal Behavior*” (Ekman, 1957) where he described difficulties in establishing empirical measurement for nonverbal behavior. Ekman, encouraged by his teacher Silvan Tomkins, reoriented his work from body movement to facial expressions. Working with Wallace V. Friesen and influenced by the work of Darwin, he conducted a series of experiments to support his universal theory. To do so, Ekman and Friesen asked people from different cultures to select emotional labels that fit the best displayed facial expressions. Surveyed cultures were from Western and Eastern literate cultures. The study has even been extended to preliterate tribes like members of the Fore linguistic-cultural³ group of New Guinea who had little exposure to literate cultures.

³“Members of the Fore linguistic-cultural group of the South East Highlands of New Guinea were studied. Until 1959, this was an isolated, Neolithic, material culture (Gajdusek, 1963; Sorenson & Gajdusek 1966). While many of these people now have had extensive contact with missionaries, government workers, traders, and United States scientists, some have had little such contact. Only subjects who had minimal opportunity to learn to imitate or recognize uniquely Western facial behaviors were requested for this experiment.” – Ekman in “*Constants across Cultures in the Face and Emotion*” (Ekman and V. Friesen, 1971)



(a) Photography of different degrees of moderate laughter and smiling.



(b) A dog in a humble and affectionate frame of mind.



(c) A Cynopithecus Niger pleased by being caressed.



(d) A Cat terrified at a dog.

Figure 3.3: Natural Expression of Emotion Across Species (Source: Darwin, 1872).

This work is depicted in the publication “*Constants across Cultures in the Face and Emotion*” (Ekman and V. Friesen, 1971). During their experiment, the procedure used by Ekman and Friesen is to present a group of three pictures of western faces simultaneously, read a story, and ask the participant to select the most appropriate picture that best fits the story. Examples of stories told are: “*his (her) friends have come, and he (she) is happy*” if the emotion to guess is **happiness**; “*his (her) child (mother) has died, and he (she) feels very sad*” for **sadness** or “*his (her) child (mother) is angry; or he (she) is angry, about to fight* for **angriness**. They concluded their experiments supporting the hypothesis that *particular* facial behaviors are universally associated with *particular* emotions. However, New Guinean failed to discriminate fear from surprise but made only a few errors when discriminating surprise from fear. The authors commented on these phenomena suggesting that not all cultures make not all the same distinction between emotions. Interestingly, they also theorized that the Fore linguistic-cultural group does not make a difference between fearful events and surprise, because fearful events are almost always also surprising (e.g. appearance of an outsider). Finally, when asked how would their face look like if they were in the situation of the different stories, members from the Fore tribesmen showed similar faces than literate cultures (see Figure 3.4).

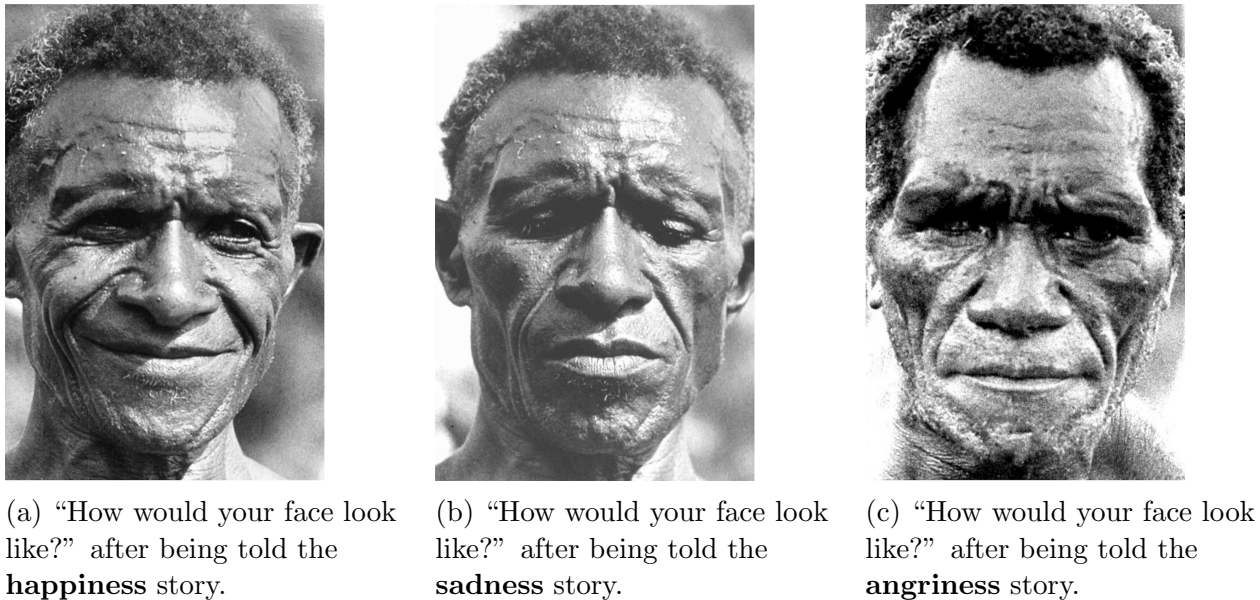


Figure 3.4: Photographies of Fore Tribesmen’ Members after Hearing Emotional Stories.⁴

These findings led Ekman and Friesen to a classification of discrete, measurable, and physiologically distinct emotions called “*basic emotions*”. Why basic? Ekman describes this term in two ways: first basics emotions are shared across all human cultures, even for preliterate tribes; secondly, these emotions are fundamental; there are not altered or produced by social learning. These emotions are listed and describes in Table 3.1.

Alongside this classification, Ekman defined a list of 13 characteristics (initially the list was composed of 11 elements in 2003, but has been extended to 13 in 2008) that are found in basics emotions. These characteristics are:

⁴source: <https://www.ekmaninternational.com/a-brief-history-into-paul-ekmans-early-research/>, (last seen 01/2020)

Anger	The response to interference with our pursuit of a goal we care about. Anger can also be triggered by someone attempting to harm us (physically or psychologically) or someone we care about. In addition to removing the obstacle, or stopping the harm, anger often involves the wish to hurt the target.
Fear	The response to the threat of harm, physical or psychological. Fear activates impulses to freeze or flee. Often fear triggers anger.
Surprise	The response to a sudden unexpected event. It is the briefest emotion.
Sadness	The response to the loss of an object or person to which you are very attached. The prototypical experience is the death of a loved child, parent, or spouse. In sadness there is resignation, but it can turn into anguish in which there is agitation and protest over the loss and then return to sadness again.
Disgust	Repulsion by the sight, smell, or taste of something; disgust may also, be provoked by people whose actions are revolting or by ideas that are, offensive.
Contempt	Feeling morally superior to another person
Happiness	Feelings that are enjoyed, that are sought by the person. There are a number of quite different enjoyable emotions, each triggered by a different event, involving a different signal and likely behavior. The evidence is not as strong for all of these as it is for the emotions listed above.

Table 3.1: Basics emotions described by Ekman (1973)

- | | |
|--|---|
| 1. Distinctive universal signals; | 8. Unbidden occurrence; |
| 2. Distinctive physiology; | 9. Distinctive thoughts, memories, and images; |
| 3. Automatic appraisal; | 10. Distinctive subjective experience; |
| 4. Distinctive universals in antecedents events; | 11. Refractory period filters information available to what supports the emotion; |
| 5. Presence in other primates; | 12. Target of emotion unconstrained; |
| 6. Capable of quick onset; | 13. The emotion can be enacted in either a constructive or destructive fashion. |
| 7. Can be of brief duration; | |

More details on these characteristics can be found in Ekman (2003) and Ekman and Lama (2008).

The collaboration between Ekman and Friesen led to the development of the **Facial Action Coding System (FACS)**, a widely used taxonomy of human facial movements. They published two major publications “*Measuring facial movement*” and the famous manual “*Facial Action Coding System: A Technique for the Measurement of Facial Movement*” in 1976 and 1977 respectively. FACS breaks down facial expressions into individual components of muscle movement, called **Action Units (AUs)**, an example is depicted in Figure 3.5. FACS are widely used in this thesis and are detailed in a further section (see Section 4.2 page 48).

Ekman and Friesen’s findings brought support to the universal theory of Charles Darwin and widely contributed to understanding the facial expressions of humans. Their publications, books, and manuals led Ekman to obtain numerous prizes and have been ranked 59th out of the 100 most

⁵Source: <https://www.paulekman.com/product/facs-manual/> (last seen 01/2020)

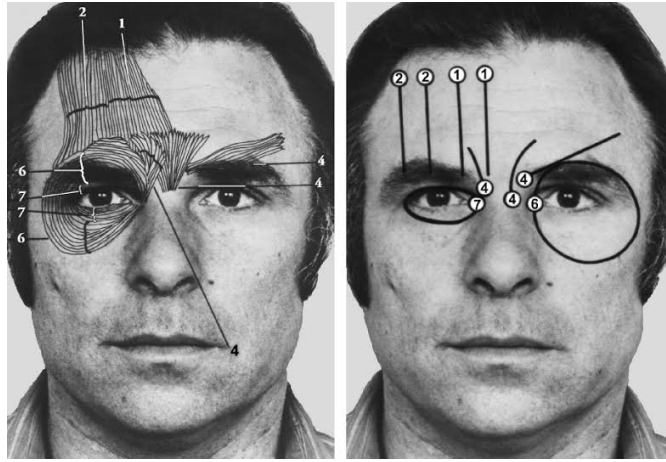


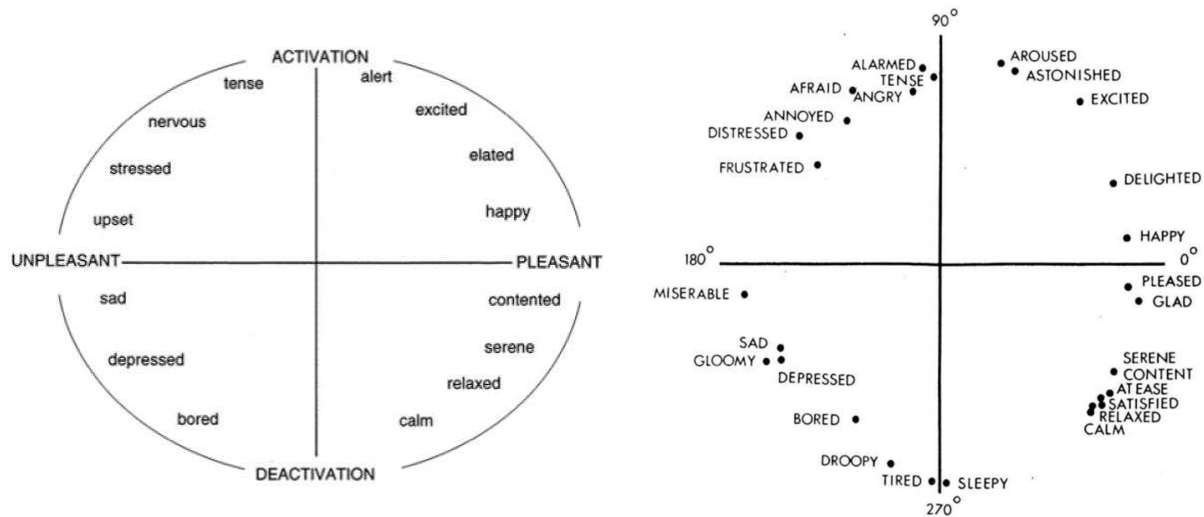
Figure 3.5: Example of the Facial Unit Coding System on Paul Ekman.⁵

cited psychologists of the twentieth century. The “basic emotion” approach inspired the community. A good example is the work from Shariff and Tracy (2011) in their interesting publication called: *What Are Emotion Expressions For?* (Shariff and Tracy, 2011). The authors proposed that facial action units had evolved in a precise goal: to communicate, in a nonverbal way, an inner mental state with other individuals. For example, if one is internally feeling fear because of an incoming threat, the corresponding facial movements immediately expressed will warn the other individuals about this threat. In the authors’ definition, humans and other animals are born with this innate ability to decode these expressions of internal mental state. Finally, they asserted that the emotional state is used to regulate the body and prepare it for the incoming situation (e.g., prepare to flee).

However this proposition of discrete emotions classification has been widely discussed and criticised among the community (Russell, 1994, 2003; Barrett, 2006, 2011). Discrete emotions classification is not the only way of representing the emotional state; other models have been proposed.

3.4 The Circumplex Model

The basic emotions theory of Ekman, presented in the previous section, has been widely used for affective computing and human-robot interaction; however, it has left numerous unsettled essential questions in the fields of psychology and psychiatry. For example, the theory does not explain patients with mood or affective disorders. Given these limitations, many theories have been developed to approach emotional states with a dimensional model. One of the first studies has been led by Harold Schlosberg in 1952, who stated that affective states are not independent of each other but were rather related according to some psychological variables. Charles Osgood supported this idea. In “*Dimensionality of the semantic space for communication via facial expressions*” (Osgood, 1966), Osgood asked student judges to label poses of student actors. A simple factor analysis yielded three principal dimensions that were similar to the dimensions proposed by Schlosberg: Pleasantness, Activation, and Control. Most theorists agree on at least two out of these three dimensions; the first one is usually named as “**Valence**” (or pleasure), which determines the pleasure intensity of the emotion ranging from extreme pain (or unhappiness) to extreme happiness. The second dimension is “**Arousal**”, also labeled as excitation or activation, ranging from sleep to medium state of drowsiness to high level of alertness. James A. Russell has proposed a similar



(a) Circumplex Model. Horizontal axis: Dis-pleasure (Left) to Pleasure (Right); Vertical axis: degree of Arousal. (b) Multidimensional scaling solution for 28 affect words.

Figure 3.6: Circumplex Model after Russell, 1980.

Affect States	Pleasure		Arousal		Dominance	
	Mean	SD	Mean	SD	Mean	SD
Happiness	0.81	0.21	0.51	0.26	0.46	0.38
Sadness	-0.63	0.23	-0.27	0.34	-0.33	0.22
Surprise	0.40	0.30	0.67	0.27	-0.13	0.38
Fear	-0.64	0.20	0.60	0.32	-0.43	0.30
Anger	-0.51	0.20	0.59	0.33	0.25	0.39
Disgust	-0.60	0.20	0.35	0.41	0.11	0.34

Table 3.2: Description of affect states using PAD model. Mean is the mean rating transformed to a -1 to +1 scale; SD is the standard deviation on the same scale (Russell and Mehrabian, 1977, p277-278).

model in *Circumplex model of affect* (Russell, 1980). It is depicted in Figure 3.6a and 28 affect words along the two axis 3.6b.

The last dimension mentioned by Schlosberg was *control*, we will call it “**Dominance**” and it represents at low-level submissiveness to dominance (or total control of the situation) at a high level. In another publication (*Evidence for a Three-Factor Theory of Emotions*, 1977), Russell and his doctoral advisor, Albert Mehrabian, claimed the necessity of this last axis, arguing that only dominance makes it possible to distinguish angry from fear, alert from surprised or relaxed from protected. For example, anger and fear become distinguishable because anger ranks high in dominance, whereas fear is submissive, placing it at the other end of the dominant-submissive axis. Finally, Table 3.2 presents the emotional state along with their corresponding coordinates in the 3D Pleasure-Arousal-Dominance (PAD) model.



Figure 3.7: Would you label the face on the left with the same emotion as the face (placed in a context) on the right? (Source: Barrett et al., 2011).

3.5 Distinction with Moods

Emotion psychologists have often discussed the difference between mood and emotion (e.g. Frijda, 2000). Generally, moods are considered as diffuse affect states, characterized by a relative enduring predominance of certain types of subjective feelings that affect the experience and behavior of a person. Moods may often emerge without apparent cause that could be clearly linked to an event or specific appraisals. They are generally of low intensity and show little response synchronization, but may last over hours or even days. Examples are being cheerful, gloomy, listless, or depressed.

3.6 Paradox and Bias in the Study of Emotion

The theories discussed in the previous sections are part of a larger pool of research on emotion and remain controversial. Russell and Barrett have proposed some interesting observations on the dominant views of emotion.

First, both Russell and Barrett questioned the validity and the relevance of Darwin's Universal theory in two publications: "*Is There Universal Recognition of Emotion From Facial Expression? A Review of the Cross-Cultural Studies*" (Russell, 1994) and "*Was Darwin wrong about Emotional Expressions?*" (Barrett, 2011). Darwin traveled across the many different cultures, distributed over many continents, in order to bring evidence for his theory. Barrett interprets this with another point of view; she asked in her publication if Darwin did not just (unconsciously) interpret these facts in a way that they fit under one unique theory (a phenomenon called *presentism*)? Is it relevant to explain all phenomena of evolution using a unique theory? Barrett also wondered if it exists alternative hypotheses that could give a better interpretation of the evidence gathered by Darwin. Neither Russell nor Barrett entirely rejects the Universal Theory, but they instead put a second thought on the premises and conclusions of this view. For example, is *happy* always expressed with raising cheeks and pulled lip corners? Barrett questioned the rigid and uncompromising association of facial muscles with emotions. In another publication (Barrett et al., 2011), she highlights how standard experiments on emotion detection are systematically conducted: the use of highly exaggerated posed faces, the display of the basic emotional words that restrict (and influence?) raters' choice and the isolation of the face. To support his view, she presented an

interesting image, presented here Figure 3.7. By analyzing the face displayed on the left, one can say that the woman is expressing severe pain⁶, but looking the whole context (right), then we understand that this is not pain, but rather a cry of relief for the victory. Barrett suggests here that information displayed on the face is not sufficient to capture the meaning of facial expression. The context plays an essential role.

An additional point concerns data annotation. Today’s research on Machine Learning requires a lot of systematic and laborious work of data annotation. To cope with this, services like Amazon’s Mechanical Turk⁷ can be used. However some recent studies highlight problems that may arise from these annotation systems (Ipeirotis, Provost, and Wang, 2010; Whitehill, Wu, Bergsma, Movellan, and Ruvolo, 2009). In the later study from Whitehill et al., the authors emphasized some theoretical and practical challenges on low-cost labeler services: a large range of unknown expertise from labelers, difficulty variation in a task and closed labels must be combined (manually) to obtain the real value. In their paper, a relevant example for our study is taken: to distinguish Duchenne⁸ smiles from non-Duchenne smiles (see Figure 3.8). This task requires knowledge about

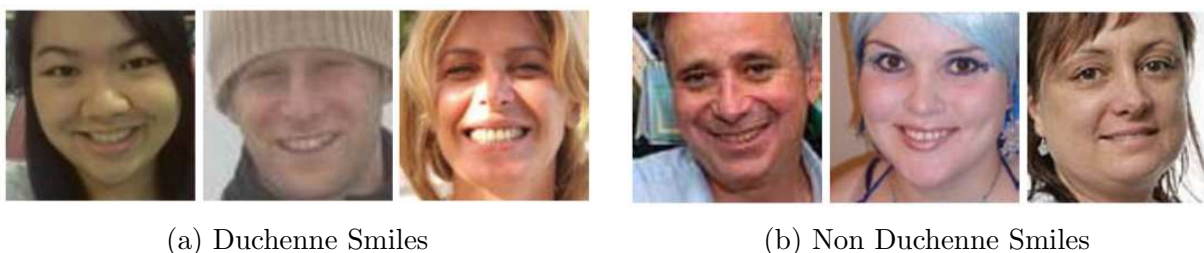


Figure 3.8: Examples of Duchenne (left) and Non-Duchenne (right) smiles. The distinction lies in the activation of muscles around the eyes, and is difficult to discriminate even for experts. (Source: Whitehill et al., 2009)

facial muscles that are involved in smiles and this could be challenging even for experts. Labelers from Amazon’s reached around 70% of accuracy; however, questions can be asked on the chosen experimental protocol here. As the Non-Duchenne smile is a social smile and labelers are familiar with the Western culture, they probably learned throughout their life of social interaction how to detect these faked smiles unconsciously. Does this mean that this ability is shared among cultures? Again context is not included in these images. A picture of a smiling woman playing with his newborn would not be labeled the same way as the same women in an uncomfortable social interaction.

The questions raised in these sections suggest that one should not accept a theory without some degree of skepticism. Without considering alternatives, experimenters may create artifacts in protocols that can influence results in many ways. If the displayed expressions by humans appear to be controversial, other researchers put their efforts analyzing the way the body is reacting biologically to emotions.

⁶I presented this image (left) to two of my coworkers. The first one concluded that the woman’s face was expressing pain or extreme sadness; the second one quickly identified the woman to be Serena Williams (a famous tennis player) and said that she should have won an important point, “*she’s happy*” he said. He knew the context, and identified the correct emotion, without looking at the image on the right.

⁷Mechanical Turks are humans that are hired to perform tasks that computers cannot do, for low-cost and with high-speed.

⁸A Duchenne smile is composed of two AUs. AU12, the zygomatic major muscle (which raises the corners of the mouth) and AU6, the orbicularis oculi muscle (which raises the cheeks and muscles around the eyes). An exaggerated smile, relying only on zygomatic muscle, is associated with a “forced” but socially accepted smile.

3.7 A Neurobiological Approach to Emotion

Based on recent findings from the Cannon-Bard theory (Cannon, 1927), the neurobiologist James Wenceslas Papez presented in 1937 theoretical and experimental data on the role of the hypothalamus, the gyrus cinguli and the hippocampus on emotion generation. A few years later, Paul Donald MacLean pursued the same idea by adding the limbic system as the central mechanism of this circuit (Papez, 1937). Through his findings on the forebrain, MacLean formulated the famous *triune brain* model in 1960 (MacLean, 1990 is the final production) which structured the brain in three different complexes: the reptilian complex, the limbic system (paleomammalian complex), and neocortex system (neomammalian complex). Over time, the research community brought evidence that some limbic structures are not directly involved in emotion generation, as Papez or MacLean suggested, but would instead participate in memory processes. However, other evidence tends to accept the role of the amygdala (Pessoa and Adolphs, 2010) and other structures in the prefrontal cortex.

Different studies contributed to the understanding of interactions and roles of both hormones and neurotransmitters in physiology regulation (and thus behavior). In 2012, Lövheim proposed an emotional state model (Lövheim, 2012) that explains the relation between the basic emotions from Ekman and neurotransmitters such as dopamine, noradrenaline, and serotonin. This model, depicted in Figure 3.9, is a cube describing relations between the secretion of a hormone (or the activation of a *monoamine* neurotransmitter) and the elicitation of a particular emotional state. According to the presented model, a basic emotion can be translated by substance generation. Table 3.3 outlined all basic emotions accordingly to the activation of monoamine neurotransmitter. For example, *Pleasure* is produced by the combination of high dopamine, high serotonin, and low noradrenaline. Finally, Lövheim emphasized the potential correlation that exists between the three neurotransmitters. Indeed, representing them on separate axis does not entail independence; they are part of a complex system of feedback that potentially interconnects all monoamine.

Such evidence has been found thanks to neuroimaging advances, using Positron emission tomography (PET) scans, Functional Magnetic Resonance Imaging (fMRI), or later real-time functional magnetic resonance imaging (rt-fMRI). These techniques' accuracy improved over time and made possible the scan of deep-brain limbic areas (e.g. amygdalae, etc.). The study of such areas helped to understand essential features for stress regulation (Gross, 2015), and the role of the amygdala was recently demonstrated in a prospective study with a priori healthy soldiers (Admon, Lubin, Stern, Rosenberg, Sela, Ben-Ami, and Hendler, 2009). In the latter study, the authors showed that stress intensity is correlated to high levels of amygdala reactivity before the stressful event.

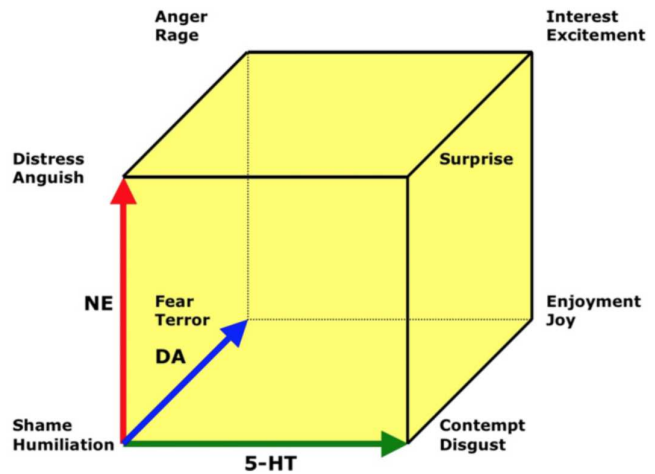


Figure 3.9: A Three-Dimensional Model for Emotions and Monoamine Neurotransmitters. The axes are: green represents Serotonin (5-HT), blue is Dopamine (DA) and red for Nor-Adrenaline (NE). (Source: Lövheim, 2012).

Basic Emotion	Serotonin	Dopamine	Noradrenaline
Happiness	High	High	Low
Sadness	?	?	?
Surprise	High	Low	High
Fear	Low	High	Low
Anger	Low	High	High
Disgust	High	Low	Low

Table 3.3: The basic emotions and assumed monoamine levels (Lövheim, 2012). Sadness is not represented in Lövheim’s cube.

Finally, the study of hormones and neurotransmitters is essential to understand emotion regulation in some psychiatric disorders. For example, the lack of serotonin secretion is related to depression (Kalia, 2005). Using Lövheim’s cube model, low-serotonin restricts emotion generation to unpleasant emotion (e.g. anger, fear, anguish, etc.), which corresponds to depressive symptoms that have been studied (inability to reach pleasure and excitement Shapsee, 2008).

3.8 Emotional Memory

We have seen in section 2.2 how memory is structured (long-term, short-term, and working memories). Perceived phenomena are stored in working memory to be processed and that older events are encoded into long-term memory for a longer period that may vary from days to an entire lifetime. These memories can be retrieved with more or less accuracy. For example, it is not uncommon to easily remember our last birthday party while it may be laborious to list again the five items that our wife told us ten minutes ago to buy at the grocery shop. How can one explain these striking variations in memory retrieval? William James observed: “*some events are so emotional as to leave a scar upon the cerebral tissues*”. It might appear, indeed, that the emotional events are more anchored than neutral ones. Brown and Kulik (1977) coined the term “Flashbulb Memory”, which described a special memory mechanism that can record almost flawlessly a particular event if it elicits emotional feelings.

Many studies have provided extensive evidence showing that this retention capacity is attributed (at least partially) to the amygdala activity. Researches from McGaugh (2000) on animals and humans defines the amygdala as the primary mechanism in emotional memory. His findings suggest that hormones and neurotransmitters activated by emotional state interact with the amygdala, which enhances the capture of the emotional event. Amygdala activities in emotional memory have been validated by numerous studies using neuroimaging (see Hamann, 2001, for a review). In particular, McGaugh mentioned the action of both noradrenergic and cholinergic in amygdala regulation.

Finally, in his review on “*Retrieval Emotional Memories*” (Buchanan, 2007) pointed out that if an emotional event has been stored in long-term memory, then exposure of a similar event (or reminder) may activate a brain activity that is comparable to the original one. This emotional state can then be directly used as a cue in for memory search.

3.9 Discussion

The study of emotion is complex. Various approaches are possible: neurobiological, face-centered, with or without context, preceding or succeeding a somatic response, and with different discrete or continuous models (basics emotions or the PAD model). Animals, including humans, express emotions to communicate non-verbally, Charles Darwin explored the world to collect evidence about this.

The study of universal facial expressions, led by Ekman and his mentor Silvan, raised awareness of Darwin’s theory of emotions. Interviews of non-literate people raised the question if emotions are cultural-dependent or innate to humans. However, as pointed by Russell and Barrett, the extensive work centered around the Universality theory influenced by the Western culture may have generated artifacts in standard experimental protocols. Including the context inside our study may be a solution.

Neurobiological models provide a new perspective on the study of emotion. Emotions can be explained with hormone levels and neurotransmitter activations. Artificial manipulation of these biological mechanisms may imply that emotion can also be artificially triggered. Nevertheless, the neurobiological approach is essential to the understanding of people with psychiatric disorders who have not accessed to some emotions (like depression).

In this thesis, we focus on the emotional expressions of chess players engaged in problem-solving. This work is not emotion-centered but will use emotion as one variable, among others. In our experiment setting, non-invasive sensors are used; thus, we can not rely on then neurobiological approach. In this study, two RGB-video based software is used. The first one is FaceReader (Den Uyl and Van Kuilenburg, 2005), which outputs basics emotions, and the second one is the state-of-the-art open-source software named as OpenFace (Baltrusaitis et al., 2018) that gives facial action units. These two softwares are described in a latter section 5.2.3.

Restrictions and controversies around basic emotions made us move from the former software to the later, less categorizing software. The emotion representation in a dimensional space is attractive and is more meaningful to explain the sudden reaction of unexpected situations. Where basic emotions include many muscle activations and intensities under a unique label, the direct use of these activation and intensity variables of action units is a down-to-earth approach.

We do not claim to offer a new approach for the study of emotion, but instead, we document and attempt to explain variations in action unit activities on the face when engaged in problem-solving. Reasoning about complex tasks requires cognitive processes, and these processes may elicit an emotional reaction that is observable. What kind of information can these observations tell us?

Chapter 4

Physiological Manifestations of Reasoning and their Observation

One of our research questions (see Section 2.5) focuses on whether it is possible to construct an instrument capable of observing external signs correlating to ongoing cognitive processes. In this chapter, we determine which external signs we are interested in this thesis. We already have some elements of answers with The Somatic Marker Theory (see Section 2.4.3) and with Facial Action Units describing the emotions described by Ekman (see Section 3.3). This chapter presents other observable signs such as gaze, described in Section 4.1, facial expressions with Section 4.2 and finally body posture with Section 4.3.

4.1 Eye-Gaze

Eye movements have been studied for centuries. When focusing on a stimulus, the eyes do not move smoothly across it. This is what the pioneer Louis Émile Javal noticed when he observed readers in 1879. He stated that readers keep alternating between short rapid movements, named as **saccades** and short stops referred as **fixations** (Émile Javal, 1879). Only direct observation was used at first, but quickly some devices based on different approaches were invented to analyze eye movements better. Understanding the cognitive process that drives fixations remains an active area of research. In this section, we review the physiology of the visual system and discuss how the visual system combines a non-linear distribution of photoreceptors in the retina with rapid eye-movements to perceive the environment, and how such motions can be tracked to reveal visual attention.

4.1.1 The Human Eye

A short introduction to the anatomy of the human eye is given in this section to understand better how eye-trackers are working.

The Anatomy of the Eye

The human eye, presented Figure 4.1, is a spherical globe that sends visual information to the brain. The **cornea** is a transparent dome that lets light enters the eye, making its way up to the pupil. To control light's intensity that goes in, a radial muscle called **iris** can enlarge or contract the size of the pupil, making the eye adaptable for different environmental illumination. Once the light has made its way through the pupil, it is first focused by a lens before reaching the **retina**,

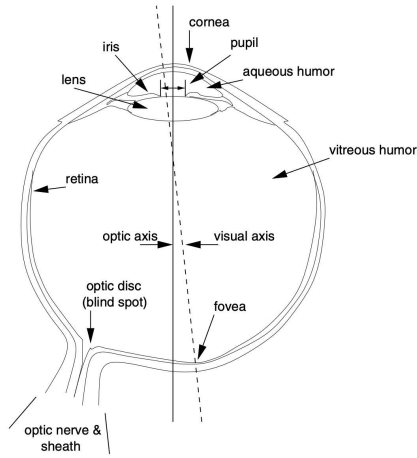


Figure 4.1: The Anatomy of the Human Eye (Source: Duchowski, 2007).

the photoreceptive layer of tissue of the eye. Besides, the human retina is a tissue composed of rods, cones, and bipolar cells. Rods and cones are responsible for human daytime and night vision, respectively.

The “Useful” Visual Field

Human visual acuity is not uniformly distributed, but highly correlated with an uneven distribution of the density of cones inside the retina. Acuity can be described in terms of projected scene dimensions (Duchowski, 2007) defined as:

$$A = 2 * \arctan \frac{S}{2D}$$

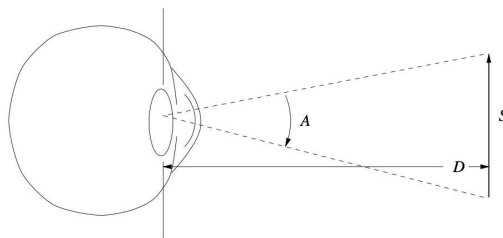


Figure 4.2: Visual Angle (Source: Duchowski, 2007).

The value of A is expressed in units of degrees visual angle, S is the size of the scene object while D is the distance to this object, see Figure 4.2. The “useful” visual field extends to about 30° , acuity drops off sharply beyond that threshold, and perceived information is mostly used for ambient motion (Duchowski, 2007). When focusing on a scene object, while the *hightest acuity* region is inside 2° , high acuity remains present at up to 4° or 5° . According to (Irwin, 1992), acuity hits only 50% of accuracy at 5° . Vision up to 6° is usually referred as *the foveal vision*, above that threshold is *parafoveal vision* (up to 30°). Visual fields are highly correlated with cone distribution density, as depicted in Figure 4.3. Indeed, studies from De Valois and De Valois (1988) found that in contrast to rods, cones provide most of the spatial information for the brain. As a result, to assure a proper resolution of the object of interest, it is necessary to move the eyes to align the pattern with the highest acuity region. Eye-tracking devices can measure these small movements.

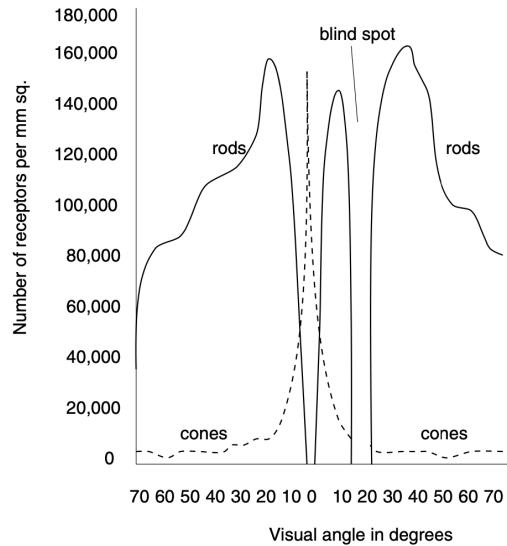


Figure 4.3: Density Distributions of Rod and Cone Receptors Across the Retinal Surface. (Source: Duchowski, 2007).

Human Visual System

The rods and cones feed signals to the optical nerve. The optical nerves leave the left and right retina via holes near the fovea to be joined at the Lateral Geniculate Nucleus (LGN) and the Superior Colliculus (SC) (see Figure 4.4). The SC acts as both a relay station to communicate retinal maps to multiple parts of the visual cortex, and as an attention filter, to suppress unattended information. The LGN provides filtered “retinal maps” to the different visual cortexes as well as to the SC. The LGN acts as a filter for visual attention, suppressing information that the system is not “attending” to (looking at). In other words, the *Superior Colliculus controls fixation* while the *Lateral Geniculate Nucleus suppresses visual stimuli during a saccade* so that the movement is not perceptible. Surprisingly, 80% of the excitation of the LGN comes from the visual cortex and other areas of the brain. The entire visual system can be seen as a succession of filters.

The output of the Superior Colliculus is a neural map that directly activates the muscles that rotate the eyes. These neural maps, or retinal maps, are directly relayed through the LGN to the primary visual cortex, where they propagate through the Dorsal and Lateral Visual pathways, see Figure 4.5. The dorsal visual pathway the sub-region of the brain colored in green while the Lateral Visual pathways are purple. The primer is called the “action pathway” and controls motor actions with spatial organization of perception, expressed as depth and direction of gaze as relayed by the Superior Colliculus. The latter (ventral visual pathway) uses color, appearance, and motion to recognize phenomena (anything that can be perceived) and objects.

These dorsal and ventral pathways are divided into several interacting subsystems (visual areas). Most human actions require input from both pathways. For example, consider the task of grasping a cup. The brain must recognize, locate the cup, and direct the hand to grasp the cup.

4.1.2 Eye-Tracking Techniques

There are two main categories of eye movement monitoring techniques: one method is focusing on measuring the relative position of the eye to the head while the second one measures the orientation of the eye in space (*i.e.* also called *point of regard*). The latter technique is extensively deployed to identify which elements are being focused in a particular visual scene (*e.g.* which chess

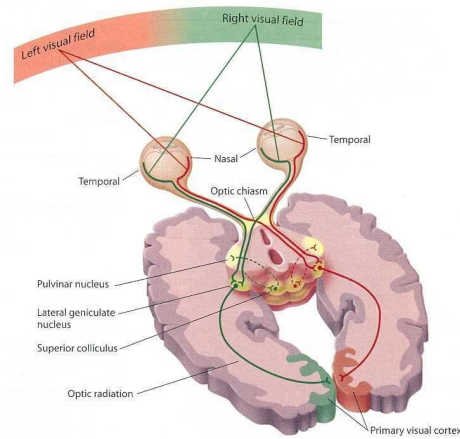


Figure 4.4: Brain Anatomy and Visual Pathways (Source: Goldstein, 1999).

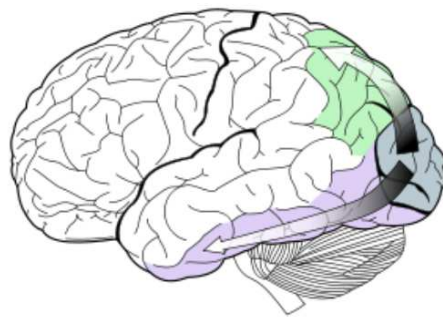


Figure 4.5: Retinal Maps Propagating Through the Dorsal Visual Pathway (green) and Ventral Visual Pathway (purple) (Source: Mishkin et al., 1983).

piece is the most focused or what elements are the most considered by customers when watching advertisements). In the past two centuries, four broad categories of eye movement measurement methodologies have been developed:

- Electro-OculoGraphy,
- Scleral contact lens,
- Photo-OculoGraphy (or Video-OculoGraphy),
- Video-based combined pupil and corneal reflection.

In the time of writing this document, the most widely applied apparatus for measurement of the point of regard is the video-based corneal reflection eye tracker, and this is the one we describe in this section.

Video-Based Combined Pupil and Corneal Reflection

By combining RGB videos from cheap and commercially available cameras with image processing hardware, video-based eye-trackers can compute the point of regard in real-time (*i.e.* orientation of the eye in space). Two setups are possible: table-mounted and head-mounted.

Table- and Head-mounted Systems Table-mounted systems, or commonly called *remote eye-trackers*, are popular among the research community as they are easily deployable and set up. Depicted in Figure 4.6 is an example of an remote eye-tracking. These are (generally) shaped as long black rectangles and are coded to operate (generally, again) when placed under a screen, facing the user (Figure 4.6a and 4.6b). The bar is composed of IR cameras and special cameras used to control illumination on users' face.. Constraints of the eye detection algorithm limit the domain of application of such remote systems: they must operate in a range between 50 to 90 cm and for screens up to 24 inches across.

Head-mounted (or *wearable*) eye-trackers are suitable for applications requiring the user to move freely. As with remote eye trackers, they employ IR cameras and light sources. However, wearable eye trackers are directly worn by the users, preferably as close as possible to the eye. They are most of the time shaped as glasses (see Figure 4.7a), with an IR camera focusing on the eye and an RDB camera recording the environment seen by the user. The mobility offered by these wearable systems extend the possible domain of application to a more naturalistic environment or outside in the real-world.

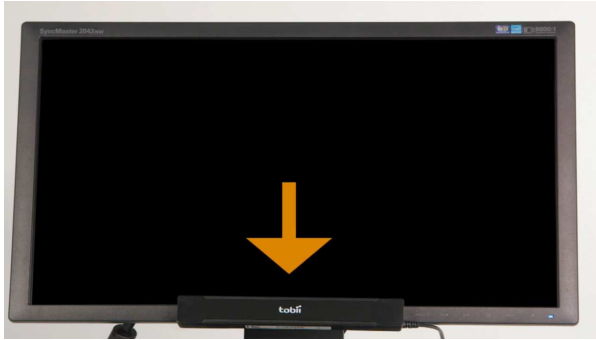
Recent growing interests in eye-tracking have made these systems much more affordable than in past decades. Examples of applications are driving assistance systems, virtual reality, or video games.

Method The idea behind the Video-Based approach is to measure the relative position of the corneal reflection of the light source (**i.e.** infra-red camera) with the pupil center (Duchowski, 2007). Due to the anatomy of the eye, four different reflections of the infra-red light are formed. These corneal reflections are called the *Purkinje reflections* (Crane and Steele, 1985).

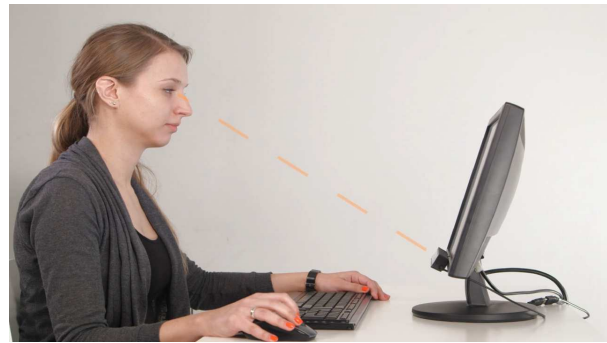
In their studies, Crane *et al.* showed that the first Purkinje reflection is relatively stable, in contrast with the eyeball (thus its pupil) that rotates in its orbit. Thus, locating the first Purkinje reflection is typically enough to estimate the point of regard. For that purpose, a calibration procedure is required before the experiment. During the calibration, the user is asked to fixate static points (generally between five and nine) while the system estimates the relative position between the first Purkinje reflection and the pupil. Figure 4.8 illustrates the procedure of calibration. Figure 4.8b shows the calibration process as seen by the user; each point has to be sequentially fixated for a few seconds.

4.1.3 Eye Movement Analysis

The goal of eye movement measurement and analysis is to gain insight into the viewer's attentive behavior. Raw eye movement data, or perhaps data processed to a certain extent such as **Gaze Intersection Point** (GIP) data in virtual reality, may appear to be informative; however, without further analysis, raw data are for the most part meaningless. Although intuitively (and from the knowledge of the task), it is possible to guess where the subject happened to be paying attention in the environment (over the internal calibration points, as she or he was instructed), it is not possible to make any further quantitative inferences about the eye movement data without further analysis. The goal of eye movement signal analysis is to characterize the signal in terms of salient eye movements, *i.e.*, saccades and fixations (and possibly smooth pursuits). Typically, the analysis task is to locate regions where the average signal changes abruptly, indicating the end of a fixation and the onset of a saccade and then again assumes a stationary characteristic indicating the beginning of a new fixation. Saccades can be thought of as signal edges in time.



(a) Table-mounted Eye-tracking, usually deployed near the interactive system, here below a computer.



(b) Relative position of the user, the interactive system (computer) and the eye-tracking. The orange dashed-line traces the eye-tracker's field of view.

Figure 4.6: A Table-Mounted Eye-Tracking System from the Company Tobii (Source: <https://www.tobii.com/>).



(a) Glasses eye-tracking device intended to be head-mounted.



(b) A driver wearing eye-tracking glasses to capture is point of regard while engaged in a driving situation.



(c) Real-time visualisation of the recorded data, the red circle displayed the point of regard current position.

Figure 4.7: A Head-Mounted Eye-Tracking (Glasses) System from the Company Tobii (Source: <https://www.tobii.com/>).

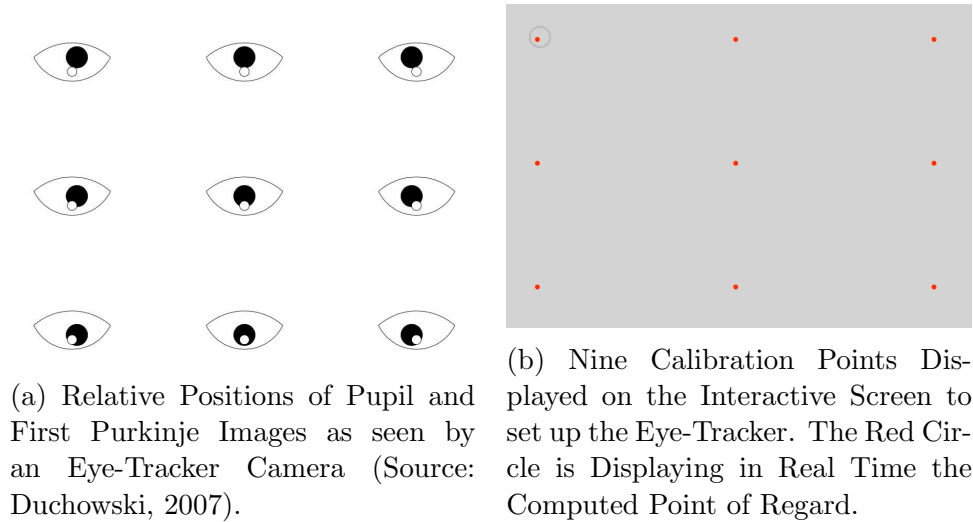


Figure 4.8: Relative Positions of Pupil and First Purkinje Images (Left) as seen by the Eye-Tracker Camera while the user is performing a Calibration (Right).

We present a list non-exhaustive of metrics, used in general Eye Movement Analysis, that are interesting for our case study.

Saccades

A saccade describes rapid eye movements from one fixation to another. Saccades typically last about 30 to 80 milliseconds (Duchowski, 2007). During this span, no information is extracted from the visual system (Holmqvist et al., 2011). The most common and used metrics related to saccades are amplitude (a distance of a saccade), duration (milliseconds), and velocity (in degrees per second).

Fixations

A fixation is an aggregation of microsaccades that focus around a center point. A single fixation's span has been determined between 200 and 300 milliseconds (Holmqvist et al., 2011). Fixations can provide interesting information about the cognitive processes involved in the task at hand. Several metrics are applicable in this respect:

Number of Fixations The number of fixations in an area of interest can give different information according to the domain of expertise, the following (non-exhaustive) list give some example that can be used in this thesis:

- **Semantic importance:** It is accepted that the general importance of a phenomenon increases the number of fixations on that particular object (Poole and Ball, 2006; Duchowski, 2007).
- **Search efficiency and difficulty:** The number of fixations on an object or AOI may reflect the ease of extracting information related to that object if few fixations have been made, whereas a large number of repeated fixations would imply difficulty in extracting salient information related to that object (Holmqvist et al., 2011). In other words, a low number of fixations could either mean that the task is too easy or that the participant is experienced.

- **Experience:** The more advanced expertise a participant has in the area being tested, the less fixation they will need to extract relevant information. This is particularly observed in the game of chess (but not only), where it has been demonstrated that experts make fewer bindings than lower level players (Charness et al., 2001).

Fixation duration Fixation duration (in milliseconds) is often associated with cognitive processing (Holmqvist et al., 2011). The general finding correlates long fixations with deep and effortful processing. This conclusion can be found in applications such as reading, scene perception, or usability. However, other findings stated that longer fixations are not always equivalent to deeper processing:

- **Longer fixations imply shallow processing:** Participant with low arousal (close to day-dreaming) may generate prolonged fixation, this is well illustrated in driving context (Chapman and Underwood, 1998).
- **Higher stress results in shorter fixation durations:** High mental workload and stress can result in more and shorter fixations (Unema and Rotting, 1990).
- **Expertise leads to longer fixation durations:** In some fields such as chess, art, or goalkeeping, experts generate longer but fewer fixations than novices (Reingold et al., 2001; Reingold and Charness, 2005). This is related, among other factors, to the improved visual system of experts who have a more elaborate parafoveal vision than novices. In these cases, longer fixations imply the extraction of information in a wider visual field. These extraction methods, therefore, result in fewer fixations and longer but more effective fixations.

Scan Path

Hölmqvist defines scan path as “*the route of oculomotor events through space within a certain timespan*” (Holmqvist et al., 2011). Alternatively, previous definitions describe scan path as a sequence of alternating saccades and fixations (Duchowski, 2007) This particular path has a start, an end, and, therefore, has a length. Several researchers agree that scan paths can be correlated with an ongoing cognitive process. However, no reliable evidence has yet been put forward to prove this hypothesis (Holmqvist et al., 2011). In order to correlate a scan path with the cognitive process, Ekhme proposed a manual analysis between scan path and verbal report (Ehmke and Wilson, 2007).

Area of Interest (AOI)

An Area of Interest (AOI) or Region of Interest (ROI) is an area defined by hand (or via software) to indicate an area of the stimulus that is particularly important for visual data analysis (Duchowski, 2007). They are usually defined with the help of experts in the field of study. In this way, the visual data in AOI will be able to receive statistical analyses such as Number of Visit in AOI (count), number of fixation of AOI (count), time before first visit in AOI (seconds), etc.

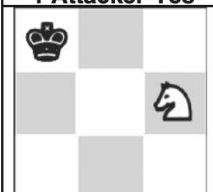


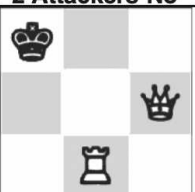
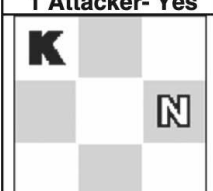
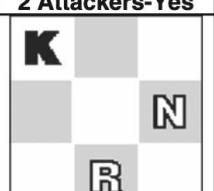
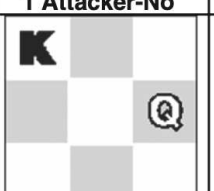
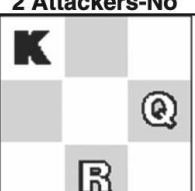
1 Attacker-Yes	2 Attackers-Yes	1 Attacker-No	2 Attackers-No
			
1 Attacker- Yes	2 Attackers-Yes	1 Attacker-No	2 Attackers-No
			

Figure 4.9: Illustration of Stimuli used in the Check Detection Task by Reingold and Charness (2005). Pieces can be either displayed with Symbols (first row) or Letters (second row). The Two First Columns Display Check Situation whereas the two last are Non-Check Situations.

4.1.4 Parafoveal Vision: The Effect of Expertise on Visual Span

We all share the same eye anatomy (except in exceptional circumstances), but does this imply that we all extract information the same way? In experiments with reading, Engbert, Longtin, and Kliegl (2002) observed that words within 1° (six to eight characters) of the point of fixation are recognized using foveal vision, while information up to 6° of visual angle can be recognized without explicit fixation. This phenomenon is referred to as the parafoveal vision. Information perceived with parafoveal vision can still be extracted by readers as long as these are not in the outer edges of the parafoveal. Common words can be perceived by readers as long as these are not beyond the outer edges of the parafoveal region. Less common and unknown words can only be extracted within the foveal region (Traxler, 2011). Thus, parafoveal vision is thought to be acquired with expertise.

The correlation of parafoveal vision with expertise is easily demonstrated in chess (Chase and Simon, 1973a,b; Reingold et al., 2001; Reingold and Charness, 2005). For example, Reingold and Charness (2005) have examined differences in the spatial distribution of fixations between experts and novices, using simple chess tasks involving two to four pieces sharing relations arranged on a 3x3 chessboard. The task, called *chess detection*, required chess players to determine if a particular relation was existing between these pieces (if the King is under attack, *i.e.* *King is in check*). The experimental procedure, described by the authors, is as follows: chess players were asked to fixate the center square board (that is always empty) before revealing the pieces. While gathering information on the board, his eye movements are monitored. When the player is ready, he can stop the task by answering the question: *is the Black King in check? Yes or no?* Example of displayed tasks are presented in Figure 4.9, pieces can be either represented by their chess symbol or by letters. Scattergrams, where dots represent an individual gaze position, are presented in Figure 4.10. A simple observation of these scattergrams reveals differences in the spatial distributions of gaze positions between skill groups. Indeed, experts present a higher concentration of individual gaze position in the center of the board compared to other levels of players. Moreover, some experts were even able to solve tasks without moving their gaze outside of the initial center square. These experts could extract information about pieces without the use of high visual acuity. These trials are named by the authors as *no-saccade trials*, as no saccade is used

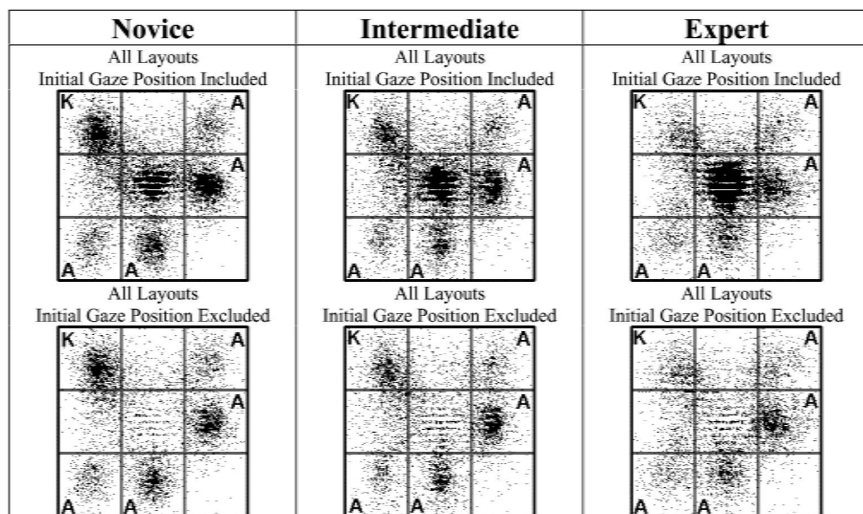


Figure 4.10: Scatter-grams of Gaze Positions in the Check-Detection Task by Skill. The first row presents data collapsed across all trial types and spatial layouts. The second row presents the same data, excluding initial gaze position. A = position of an attacker piece; K = position of the King. (Source: Reingold and Charness, 2005).

to exit the center position to fixate pieces or other squares. The proportion of such phenomena is displayed in the bar plots 4.11 for each skill groups.

However, only experts were able to demonstrate this ability. This ability suggests that experts benefit from a larger visual span for chess-related visual patterns. Another argument supporting this conclusion is the number of fixations made by experts (fixations in *no-saccades trials* are not considered), these results are depicted in Figure 4.12a and 4.12b. Experts made fewer fixations in general and, when fixations are recorded, they are proportionally less positioned on the pieces than for lower-level players. In conclusion, Reingold *et al.* demonstrated that the increase in the visual span is a function of skill. Experts can extract more information from fewer fixations that cover a more significant portion of the chessboard than for intermediates or novices players. In other words, they can extract more information from their *useful visual field*, which is not only about the highest acuity region (the foveal vision) but also about more open angles of their parafoveal vision. This conclusion is consistent with the hypothesis from Chase and Simon's (Chase and Simon, 1973a) stating the potential increase in visual span for experts is related to an encoding advantage learned through experience.

Finally, based on these findings, Reingold *et al.* (Reingold et al., 2001; Reingold and Charness, 2005) stated that experts are not advantaged in quick identification of singles chess pieces and board locations, but rather in the extraction of meaningful relational information that bound pieces together on the board. This main perceptual advantage is directly linked to the formation of meaningful chunks in an expert's memory, which allows them to efficiently and automatically extract several chess relations in a short time. This strong statement is demonstrated by, on the one hand, the solid performances of experts in tasks where pieces are bounded by relational and meaningful information and, on the other hand, their total absence of skill with a random configuration of pieces.

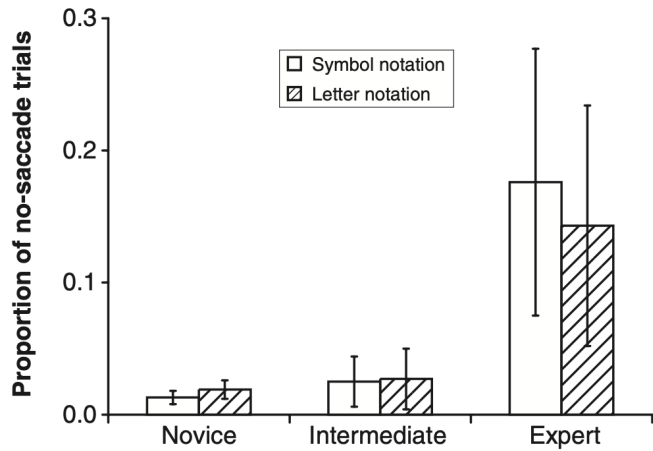
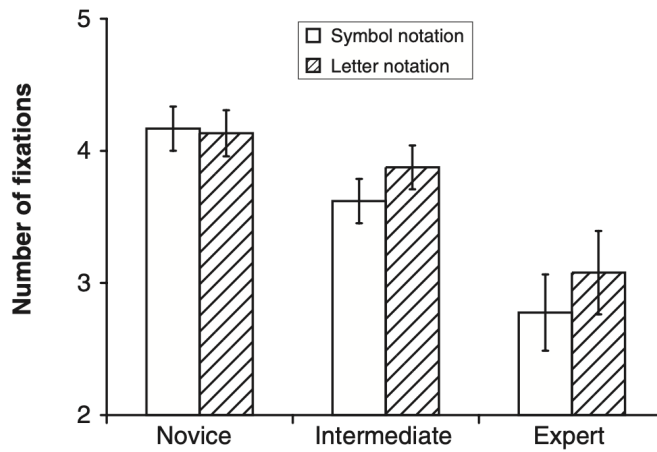
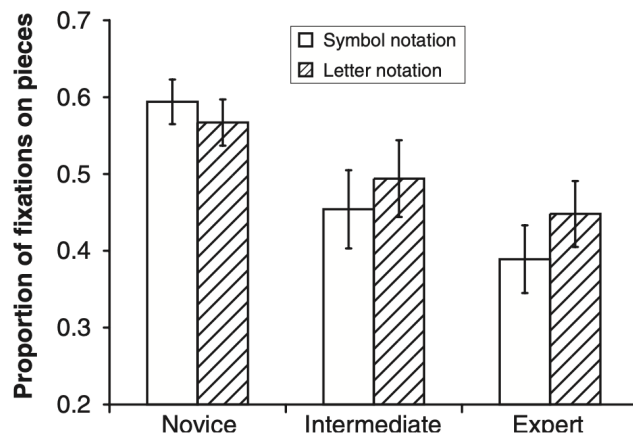


Figure 4.11: Proportion of No-Saccades Trials (Source: Reingold and Charness, 2005).



(a) Number of Fixations



(b) Proportion on Fixation on pieces

Figure 4.12: Proportion on Fixation on the board (Top) and only on Pieces (Bottom) (Source: Reingold and Charness, 2005).

4.2 Facial Expressions

We have seen in Section 2.4.3 that during human reasoning or decision making, both conscious and unconscious neural operations are involved (Damasio, 1996). During these neural operations, somatic signals (or markers) are observable on the individual: variation in heart rate, facial expressions, pupil dilation, body postures, muscle contractions, etc.

4.2.1 Facial Action Coding System (FACS)

Section 3.3 discussed Ekman’s theory that all humans share six categories of basic emotions. This has led to the development of a taxonomy of human facial movements, called the Facial Action Coding System (FACS) (Ekman, 1977).

Ekman and Friesen’s studies led to the development of a taxonomy of human facial movements, called the **Facial Action Coding System** (FACS). This taxonomy is based on atomic element called Action Units (AU), see Figure 4.13.

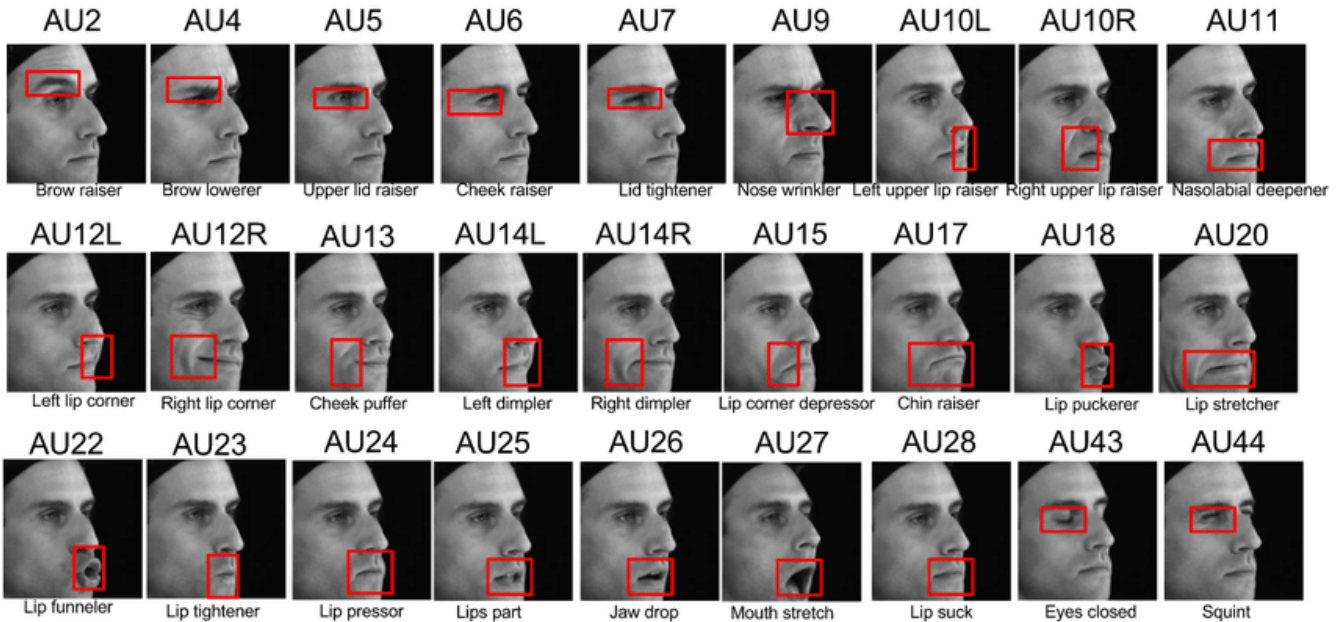


Figure 4.13: Subsets of AUs with Fully Activated Muscles. AU12L and AU12R are Distinct; Similar for AU14. (Source: Xiang and Tran, 2017)

AUs are fundamental actions of individual muscles or groups of muscles. FACS gave the scientific community a tool for facial recognition as well as facial modeling; the medical community welcomed this system in psychology where psychotherapists could use it to understand how their patients’ face behaved, and animators enjoyed this new tool to animate more realistic facial movements. Last but not least, the most interesting for our work is that each basic emotions can be decomposed into a group of muscles that can be easily recognized with FACS. Indeed, in 1983, the same authors described a system Emotional Facial Action Coding System (EMFACS) where they consider only emotion-related AU (Friesen et al., 1983). Considering that, in FACS, AU are enumerated from 1 to 46, basic emotions can be recognized by detecting the corresponding AU as described in the Table 4.1. For example, to recognize Happiness, one has to detect on the subject face if the Action Unit number six (AU6), which corresponds to “Cheek raiser”, and the AU12 (Lip corner puller) are both activated by muscle contraction.

Emotions	Action Units
Happiness	6 + 12
Sadness	1 + 4 + 15
Surprise	1 + 2 + 5B + 26
Fear	1 + 2 + 4 + 5 + 7 + 20 + 26
Anger	4 + 5 + 7 + 23
Disgust	9 + 15 + 16
Contempt	R12A + R14A

Table 4.1: Action Units associated with basic emotions according to the EMFACS.

Note: Numbers are AUs, letters are of two kinds: ‘R’ or ‘L’ means on the “Right” side or “Left” of the face respectively; ‘A’ to ‘E’ is the intensity score of the muscle contraction, ‘A’ being the minimal (trace) and ‘E’ the maximum level of contraction.

4.2.2 Action Units (AU) Detection Systems Pipeline

In modern systems for AU detection, the most common pipeline can be summarized as follow:

- From an input image, perform a Face Detection;
- From a small region containing a face, perform a 3D Facial Landmark detection (Facial landmarks are standard reference points that can be easily identified. Such as left and right junctions of the upper and lower lips, left and right junctions of the upper and lower eyelids, eye pupils, the tip of the nose, etc.);
- From 3D landmarks, estimate the head pose and perform a face alignment;
- From an aligned face, detect Action Units (AUs) and their activation intensity between 0 and 1;
- Additionally, extract other information (e.g. eye gaze orientation).

An example of pipeline used by the software Openface (Baltrusaitis et al., 2018) is depicted in Figure 4.14.

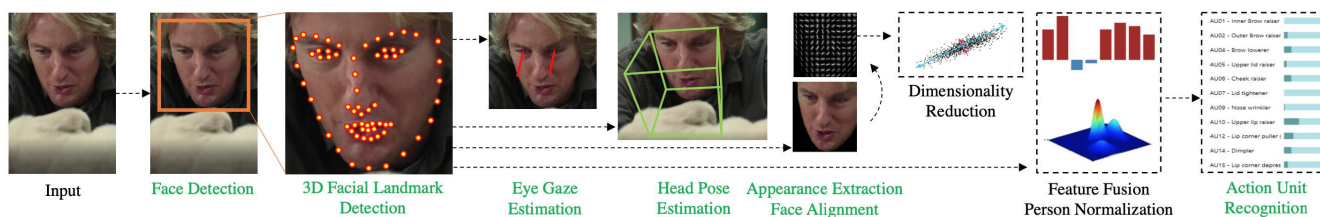


Figure 4.14: OpenFace 2.0 Facial Behavior Analysis Pipeline (Source: Baltrusaitis et al., 2018).

4.2.3 From Actions Units to the Pleasure-Arousal-Dominance Model

Strict categorization of facial action units into basics emotion can result in the loss of information about variation in the intensity of muscle activations inside a category. In some application domains, it is more interesting to observe relative variations of muscle activation than the emotion

category itself. One way to ensure that is to map Facial Action Units into the PAD (Pleasure-Arousal-Dominance) model proposed by Russel (Russell, 1980). We present here the calculations used to determine values for **Valence** (Pleasure) and **Arousal** from AUs activations. These calculations are based on publications describing FACS (Ekman, 1977) and documentations provided by Noldus FaceReader Software (Technology, 2015).

Pleasure (or Valence)

The emotional state (**i.e.** being pleased or dissatisfied about a situation or something) is given by the Pleasure indicator or commonly called **Valence**. It is computed using the activation intensity of AUs involved in positive emotion from which we subtract activation intensity of negative emotion AUs. AUs involved in positive affect are: 6 and 12. Those classified as negative are: 1, 2, 4, 5, 7, 9, 15, 16, 20 and 26. Let $AI_n \in [0, 1]$ be the activation intensity of the Action Unit n (AUn). POS_{aus} and NEG_{aus} are the normalized values for both categories of positive AUs and negative AUs respectively, such as:

$$\begin{aligned} \{P\} &= [AI_6, AI_{12}], \\ \{N\} &= [AI_1, AI_2, AI_4, AI_5, AI_7, AI_9, AI_{15}, AI_{16}, AI_{20}, AI_{26}], \end{aligned}$$

$$POS_{aus} = \frac{1}{\#\{P\}} \sum_{i \in \{P\}} AI_i \quad (4.1)$$

$$NEG_{aus} = \frac{1}{\#\{N\}} \sum_{j \in \{N\}} AI_j \quad (4.2)$$

Thus, the valence value, V , can be obtained by subtracting the positive activations from negative's:

$$V = POS_{aus} - NEG_{aus} \quad (4.3)$$

$V \in [-1, 1]$ where -1 is very dissatisfied, 0 represents a neutral emotional state and 1 a very pleased expression.

Arousal

Excitation, represented with the indicator **Arousal**, indicates whether or not the individual is active or not. Arousal value is computed following four steps:

1. Take as input the activation intensity AI_n of 20 AUs:

$$\{A\} = [1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 18, 20, 23, 24, 25, 26, 27, 1 - 43] \quad (4.4)$$

The opposite value of AU43 (eyes closed) needs to be taken as its activation indicates non-excitation.

2. For each AI_n , compute its mean value, AAI_n over the last 60 seconds, starting from $t - 1$:

$$AAI_n = \frac{1}{60} \sum_{i \in [t-61, t-1]} AAI_{n,i} \quad (4.5)$$

3. Apply a correction to avoid bias from continuously activated AU using the current activation value at time t_0 :

$$CAI_n = \text{Max}(0, AI_{n,t_0} - AAI_n) \quad (4.6)$$

4. Finally, the **Arousal** value A is calculated using the mean of the five highest values found in equation 4.6:

$$A = \frac{1}{5} \sum_{i \in [1,5]} (i^{\text{th}} \text{ max value of } CAI_n) \quad (4.7)$$

$A \in [0, 1]$ where 0 is not excited as all (**e.g.** very quite, or even sleepy) while 1 is very aroused. Note that A is often mapped between $[-1, 1]$ to be in the same range as Valence.

Dominance

Dominance models the degree to which a subject feels in control of a situation or perceived phenomena. To the best of our knowledge, there is no literature on how to determine **Dominance** using (only) Facial Action Units. Dominance tends to be perceived as a continuous state involving body posture, speech, gesture and cognitive state (see next Section 4.3) (Mehrabian and Friar, 1969; Argyle, 1988; Grammer, Fink, Oberzaucher, Atzmüller, Blantar, and Mitteroecker, 2004; Johal, Pellier, Adam, Fiorino, and Pesty, 2015). For example, as described by Argyle (1988), standing with an expanded chest and hands placed on hips is generally perceived as a dominant posture. Whereas Mehrabian and Friar (1969) stated that posture communicates dominance through relaxation, displaying, for example, asymmetrical arm and leg positions, hands relaxed and backward lean.

In this thesis, we do not compute a value for **Dominance** as no consensus has been found yet among the scientific community. However, we can still interpret the results we obtained using literature.

4.2.4 Cognitive Interpretations

The role of the prefrontal cortex (PFC) has been widely studied to understand its role in decision making (Bechara et al., 1994; Fuster, 1997). In his research, Damasio found that individuals presenting damages in their PFC were insensitive to the future consequences of their choices (Bechara et al., 1994). The authors concluded the significant role of PFC in weighing future consequences of affectively salient decisions. Fuster stated that PFC “*integrates, organizes, and structures the primitive sensations of pleasure and arousal with knowledge of the temporal contingencies that link prior experiences of stimuli within varying life contexts with expectations for the future*” (Fuster, 1997). For example, the experience of high pleasure and high arousal from winning a lottery is partly the result of the present context, previous financial difficulties, and expectations of future advantage, or both. Assimilating this information into the PFC results in an experience of conscious joy. Finally, Russel summarized the same view by stating that the cognitive functions of PFC have an essential role in the creation and recognition of emotions. This is done by association and integration of neurophysiological sensations with internal (from memories and current thoughts) and external cues (Posner, Russell, and Peterson, 2005).

4.3 Body Posture

Darwin questioned the purpose of different body reactions: Why are we displaying disgust with wrinkling nose? Does baring teeth when enraged is making sense even if biting has a low probability of occurring? Darwin approached these questions from an evolutionary point of view, stating that these body reactions were vestiges from old behaviors that had specific functions (Darwin, 1872). Even if most of these functions disappeared today (biting is rare), these reactions are still used as external evidence to communicate the individual's internal state. We still wrinkle nose to communicate disgust, even if the stimuli is an odorless picture. This "stimulus-response" interpretation, derived from behaviorism theories, has limits and does not take into account cognitive processes such as emotion, memory, thinking, or problem-solving (Miller, 1956; Damasio, 1991). Mehrabian and Friar stated that changes in the individual's affective state are reflected in body language (Mehrabian and Friar, 1969). Ekman and Friesen, on their side, refers as unintended "emotional leakage" when interpreting somebody behaviors (Harrigan, 1985; Ekman and Friesen, 1969). Damasio, in his Somatic markers theory, assumed that neural operations such as human reasoning and decision making are reflected through external somatic signs (e.g. variation in heart rate, pupil dilation, etc.) (Damasio, 1991).

Over the last decades, researches have been studying non-verbal behavior through five significant areas: facial expression, nonverbal vocal behavior, kinesics, visual behavior, and proxemics (Harper, Wiens, and Matarazzo, 1978). This section focuses on cognitive processes that can be elicited and observable through body behaviors in a chess-playing environment: reactions to stress, self-touches, and external signs of dominance.

4.3.1 Physiological Responses to Stress

Definition and approaches

The concept of stress has been studied for decades (Selye, 1950), with proposals for a variety of definitions and approaches (Koolhaas, Bartolomucci, Buwalda, de Boer, Flügge, Korte, Meerlo, Murison, Olivier, Palanza, et al., 2011). In their critical evaluation of the stress concept, Koolhaas *et al.* supported a "cognitive perception" approach to define stress: "a cognitive perception of uncontrollability and/or unpredictability that is expressed in a physiological and behavioural response" (Koolhaas et al., 2011). They emphasized the non-reversibility of their definition, and observable physiological responses do not necessarily imply stress.

This complex state, involving processes interacting with each other, needs to be studied from a more holistic point of view. With this in mind, Aigrain and colleagues (Aigrain, Spodenkiewicz, Dubuiss, Detyniecki, Cohen, and Chetouani, 2016a) analyzed stress in 3 perspectives: the biological perspective, the phenomenological perspective, and the behavioral perspective:

- The **biological** perspective focuses on how the body reacts to stressful stimuli. Reactions by the hypothalamo-pituitary-adrenal pathway and the autonomic nervous system trigger a neuroendocrine chain reaction, resulting in the release of the hormones epinephrine and cortisol from adrenal glands (Selye, 1976). This, in turn, affects other systems such as cardiovascular, gastrointestinal, or the nervous systems.
- The **phenomenological** perspective considers the individual's perception of stress. This view has been mainly put forward by Lazarus' theory of cognitive appraisal (Lazarus, 1966, 1993) stating that stress is a two-way process between the stimuli and the individual ability to gather resources required to manage both the situation and the produced stress. In this

view, stress would occur only when the individual perceives that its abilities or resources to cope with the stressful situation are not enough.

- The **behavioral** perspective investigates the impact of stress on human and animal behavior both at individual and group levels. An example is displacement behaviors (**e.g.** scratching, face touching, and lip biting) observed in humans and primates (Troisi, 2002). These body gestures carry more information about the emotional state of individuals involved in a stressful situation than other external cues such as verbal expressions. These behaviors have been interpreted as short diversions used to cut-off attention temporarily from the stressful stimuli, discharging the excess energy accumulated to face the situation. However, these diversions could result in reducing the individual’s abilities (Mohiyeddini and Semple, 2012; Weisman, Chetouani, Saint-georges, Bourvis, Delaherche, Zagoory-Sharon, and Cohen, 2015).

Stress detection

In their work of multimodal stress detection, Aigrain *et al.* (Aigrain et al., 2016a) reviewed numerous automatic stress detection systems that involve different stressing stimuli, signals, and annotations. For example, Wijsman *et al.* (Wijsman, Grundlehner, Liu, Hermens, and Penders, 2011) asks individuals to realize mental arithmetic, a logical puzzle, and memory tasks while being monitored their heart rate, electrocardiography, electromyogram, and respiration. For this study, stress annotation has been performed from a self-assessment. Aigrain *et al.* summarized all observable signals that have been used in these systems:

- **Physiological** signals: Blood Volume Pressure (BVP), Electrocardiography (ECG), Electromyogram (EMG), Galvanic Skin Response (GSR), Heart Rate (HR), Heart Rate Variability (HRV), etc.
- **Behavioral** signals: speech, body movement, head position, etc.

To tackle the complexity of stress detection and the high dimensionality problem, Aigrain *et al.* proposed to evaluate the predictive power of 101 behavioral and physiological different features. These features are extracted from:

- Body features (**e.g.**, head position and posture enhance),
- Quantity of Movement (**e.g.**, displacements of the skeleton joints),
- Number of time the body is in periods of high activity,
- Posture changes (**e.g.**, arm crossing),
- Self-touching (**e.g.** face touching, rubbing fingers),
- Facial features (**i.e.** 12 AUs)
- Physiological features (**e.g.** cardiac functions, respiratory system or electromyographic activity).

They observed these 101 signals produced by individuals engaged in a mental arithmetic task. In a second time, classification performances of these features are evaluated individually and in different subsets. Authors concluded that features related to body movement, blood volume pulse, and heart-rate provide valuable information for the stress detection classification system.

Drawback on stress assessment

Caution is required when evaluating stress. Methods for automatic detection of stress no matter how reliable, depend on the way stress is assessed. In their work, Aigrain et al. (2016a) reviewed the variety of stress assessment:

- Self-assessment;
- Assessment from biomarkers (cortisol, GSR, etc.);
- Assessment from external observers;
- Inference from experimental conditions;
- Inference from acting instructions.

The authors stated that most of the time, other studies on stress detection rely on only one of these assessments. Besides, Lutchyn *et al.* suggested that inconsistent results are partially explained by the choice of assessment that evaluates stress on only one spectrum (e.g. behavioral or physiological) (Lutchyn, Johns, Czerwinski, Iqbal, Mark, and Sano, 2015).

Based on these observations, the authors proposed a multi-assessment methodology to analyze stress relying on three different stress annotations:

- External Observer Assessments using the crowd-sourcing platform (e.g. behavioral perspective),
- Self-Assessment (e.g. phenomenological perspective),
- Physiology Expert Assessment (e.g. biological perspective).

When applying their 101 features (described in the previous paragraph), they provide more robust results that avoid frequent inconsistency found in other studies (Aigrain et al., 2016a).

4.3.2 Self-Touching

Nonverbal channels are thought to express emotion, attitude and psychodynamics attributes (Harper et al., 1978; Harrigan, 1985). In particular, self-touching has been interpreted as an indicator of affective state (Ekman and Friesen, 1969). Self-touching is defined as placing a hand on one’s body, followed by scratching, grooming, or rubbing. These gestures are generally executed with awareness. However, studies have shown that these behaviors occur with regularity and highly situation-dependent. Indeed, self-touches are automatic responses to bodily needs. Ekman and Friesen suggest in (Ekman and Friesen, 1969) that these behaviors are “*emotional leakage*” reflecting one’s current affective state and they are unintended realized to assure “*survival of the organism*”. More recently, these acts have been categorized as “displacements behaviors” (Troisi, 2002) that are short diversions to the brain used to cut-off attention temporarily from the stressful situation. This view is in line with Freedman’s psycholinguistic perspective (Freedman, 1972), in which self-touches occur when an individual enters in “*conflict between what remains un verbalized and is expressed only in motor form, and what is verbalized*”.

In his study, named “Self-Touching as an Indicator of Underlying Affect and Language Processes”, Harrigan studied the impact of self-touching by doctors and patients during medical interviews (Harrigan, 1985). He found significant results that support previous studies, self-touching can be explained by:

- The **Affect** theory: these gestures are indicators of stress, tension, and negative affects;
- The **Psycholinguistic** theory: “*self-touching is coordinated with information production or processing*”.

Harrigan concluded on emphasizing that self-touching should not always attribute to negative affect, but should rather be seen as a complex reaction from the body.

4.3.3 Manifestations of Dominance

Mehrabian *et al.* stated that changes in body postures reflect individual’s affective states (Mehrabian and Friar, 1969). Following this idea, cues of individual dominance or submission should be observable in body behaviors. Argyle (1988) described a dominance posture with expanded chest while being standing and hands placed on hips. Relaxation is also associated with dominance and power in an established hierarchy. This is proposed by Mehrabian, who defined dominance through relaxation (Mehrabian and Friar, 1969).

Johal tested different styles of parenting behavior for robots, and found that subjects were able to detect dominance and authoritativeness from robot posture (Johal, 2015). In this view, they reviewed works from Mehrabian and Friar (1969) to establish a list of body gestures that reflect **Relaxation** which is a concept, tossed by Mehrabian, associated with dominance and power in an established hierarchy:

- Arm-Position Asymmetry;
- Hand Relaxation;
- Sideways Lean;
- Neck Relaxation;
- Leg-Position Asymmetry;
- Reclining Angle.

Most of these cues are studies in a environment where two (or more) individuals are interacting. Johal *et al.* observed that postural openness is significantly perceived as a sign of relaxation and dominance by individuals. More specifically, individuals tend to occupy less volume with their bodies when they perceived themselves in a state of submission.

4.4 Conclusion

This chapter provides information on the physiological manifestations of reasoning and how to observe them.

Gaze analysis has been studied for several decades, and a variety of hardware is now available. The chess game has been used many times by researchers to conduct studies on the attention, expertise and visual system of players (De Groot, 1978; Charness, 1981; Reingold et al., 2001; Reingold and Charness, 2005). With these studies, we have relevant baselines and reliable results on which to compare our results. The chessboard, by design, can be easily divided to create AOI and conduct our studies. Moreover, eye-trackers offer relevant metrics for our work: the bindings reflect the extraction of information, and the scan paths allow us to identify the observation paths used by the players (Duchowski, 2007; Holmqvist et al., 2011).

Secondly, facial expressions are very easily observable with a simple camera and using the FACS description defined by Ekman (1977). From these data, we can obtain the emotional reactions of chess players and represent them in several forms: activation of micro-expressions, discrete

classification with basic emotions (Ekman, 1977) or continuous classification with the PAD model (Russell, 1980).

Finally, body behaviors also reveal relevant information about the mental state of humans. Stress can be observed in several ways (Aigrain et al., 2016b). Self-touches may appear as conscious and meaningless movements, but studies have shown that they reveal information about a mental state overloaded by stress or by too much expensive cognitive work (Harrigan, 1985). Finally, the general posture of the body reflects a state of dominance or submission that can be observed according to several criteria (Argyle, 1988)

Detecting and correlating these manifestations of reasoning could allow us to understand the mental state of a player engaged in problem-solving. To detect these external signs, we have developed a recording instrument, composed of several hardware components, and we have tested it with a first experiment, described in the following chapter.

Chapter 5

Observing Human Reasoning during Problem-Solving

5.1 Previous work

The extensive work done by De Groot and his colleagues in the search for the difference between chess Masters and Novices raised much interest among the communities. Among other interesting findings, De Groot (1978) sketched out four different phases used by players when presenting a new chess task. The first phase is perception-oriented, subject gathers as much meaningful information as possible and is not concerned (yet) about seeking a solution to the problem. In 1969, Simon and Barenfeld analyzed this particular phase and proposed a program that could imitate the information processing of a player. In two attempts, Simon and his colleagues elaborated two systems: PERCEIVER (1969) and MAPP (1973). The former simulates the eye scan that a subject would make when facing a new chess problem; while the latter tries to reproduce how humans scan a chessboard for a reconstruction memory task.

5.1.1 The PERCEIVER System (1969)

Herbert A. Simon and Michael Barenfeld proposed that problem-solving should not be limited as only selective search through a “tree” of solution possibilities, but as only a part of a more critical process (Simon and Barenfeld, 1969). They observed that search theories failed to describe the perceptual behavior of subjects when presented with a new task to solve. Those first seconds of exposure, described as a “perceptual” phase by de Groot, would require different cognitive processes than those involved in the conventional search algorithm (Tikhomirov and Poznyanskaya, 1966). In their paper, Simon and Barenfeld first describe a method that relies on eye-tracking records and then sketched out a single computer program that reproduces human behavior during the first phase of exploration.

Early Systems on Eye Movements Tracking in Chess

Strategies used by chess players to solve a task during the first seconds of exposure have been studied by De Groot (1978). While he initially relied on verbal reports, the need to observe the sequence of visual fixations of chess players soon became essential to progress. Extensive work had been performed in order to obtain satisfactory accuracy for eye movement recording (Tikhomirov and Poznyanskaya, 1966; De Groot, 1978). At that time, a fixation position could be evaluated within a visual span with a size barely larger than a chessboard’s square. However, due to the size

of a standard chessboard, the distribution of the pieces and the eye's ability to gather information using Peripheral Vision (*i.e.* the vision outside the fixation point, using parafoveal vision), Simon and Barenfeld pointed to a recurring problem in eye-tracking systems: *"Records of eye movements can only show the succession of fixations; they cannot show precisely what information is being processed at each moment"*.

The PERCEIVER System

Simon and Barenfeld designed the chess-program PERCEIVER (1969) that simulates the eye movements of humans during the perceptual phase. Based on the findings from De Groot (1978), who stated that players gather information by perceiving the relations between pieces, Simon and Barenfeld established two assumptions to describe PERCEIVER:

1. During the perception phase, information gathered is the relation between pieces. Relations are:
 - (a) Defending: a piece defending another one;
 - (b) Attacking: a piece attacking another one;
 - (c) Defended: a piece being defended by another one;
 - (d) Attacked: a piece being attacked by another one.

Chess regroups more relations between pieces, but PERCEIVER only used these four.

2. When a piece A is fixated and if A shares a relation with another piece B, than the next fixation is either on B or on A once B has been quickly checked (by peripheral vision or short saccade).

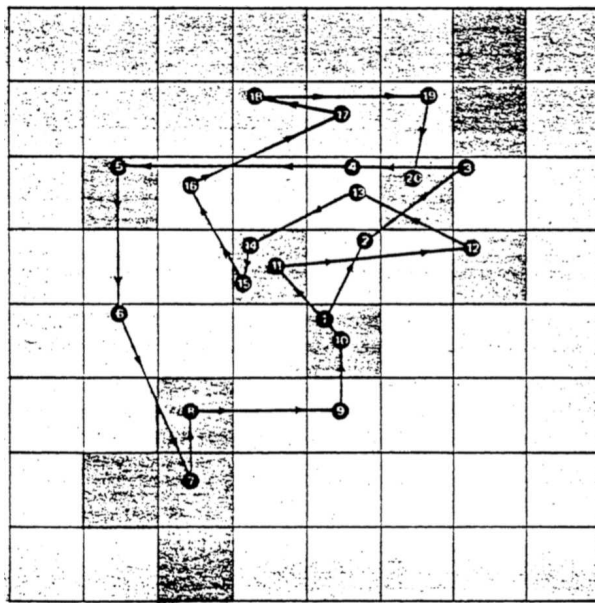
Based on these assumptions, Simon and Barenfeld compared eye movements recorded from an expert player and a simulation of PERCEIVER. For that, they decided to use a game that was already used by Tichomirov and Poznyanskaya in their 1966 study and is presented in Figure 5.1. This game was presented for 5 seconds to the expert player and PERCEIVER during his phase of exploration. A total of 20 fixations has been recorded for the expert player, while the chess-program PERCEIVER generated 15.

Results, depicted in Figure 5.2, show that both the human player and the chess-program are concerned with pieces that have an impact on the center, *i.e.* the two facing Pawns, Black and White Knights and Queens. A significant remark raised by Simon and Barenfeld is that PERCEIVER always fixates squares occupied by pieces due to its conception. On the contrary, the expert sometimes seems to fixate on empty squares. Simon and Barenfeld explained that this phenomenon could be caused by some errors in the eye-tracking calibration or other unknown reasons. It has been proved in latter study (Reingold et al., 2001) that experts can gather information from peripheral vision, which is drastically larger than novices. By looking into details which relation inter-pieces PERCEIVER noticed, Simon and Barenfeld concluded that the chess-program was concerned by the undefended Pawn and found four different moves to overcome the situation. The same solution has been considered by expert players. From this first experiment, Simon and Barenfeld concluded that a perceptual-based chess-program could be elaborated to simulate human eye movements in an exploration phase.

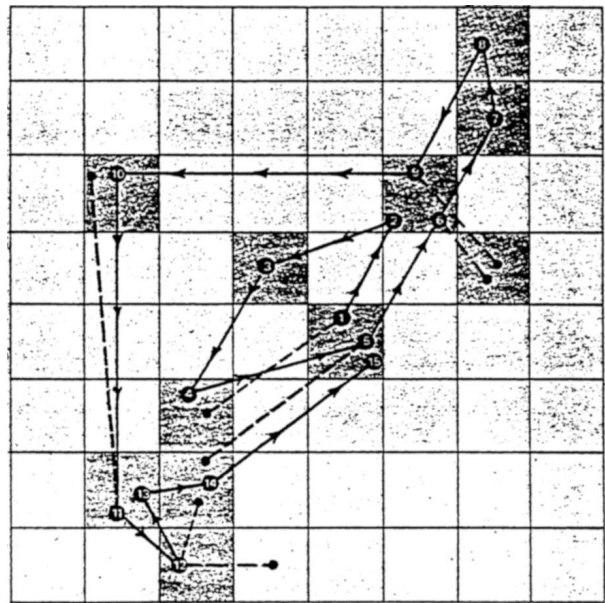
To deepen their study, Simon and Gilmartin explore encoding and retention of information. They turned to an information-processing theory component theory known as EPAM (Elementary Perceiver And Memorizer) from Gregg & Simon, 1967. This particular component focuses on



Figure 5.1: Task Presented to both an Expert Player and the PERCEIVER (Black to play).



(a) Recording of an Expert's Eye Movements during the First 5 Seconds of Presentation of the Task.



(b) Simulated Eye Movements from PERCEIVER Program. Solid Lines are Eye Movements while Broken Lines describes Relations between Pieces Noticed.

Figure 5.2: Recorded Eye Movements of an Expert Player (left) and the PERCEIVER Program (right) on Task Presented in Figure 5.1. Shaded Squares are those where Relevant Pieces are Standing. (Source: Simon and Barenfeld, 1969).

simulating recognition and fixation processes in learning. Encoding and decoding information is also part of EPAM. Learned configurations by EPAM are stored as a discrimination net that can be used when a stimulus is presented. EPAM is able to retrieve a configuration from several stimuli composed of chess pieces and store this information in what could represent the short-term memory. They concluded that this combination of two information-processing theory components could simulate the recognition and retention processes of subjects engaged in memory tasks. This is described in the next section.

5.1.2 The MAPP System (1973)

Simon and Gilmarin (1973) developed and simplified the combination of the PERCEIVER and EPAM modules to simulate how humans reconstruct the chessboard after only brief exposure. They proposed a new chess-program called MAPP for Memory Aided Pattern Perceiver. It is composed of two main components: a Learning and a Performance modules. The learning component is a derivate of the EPAM module, which simulates the storage of information in long-term memory. The amount of information stored can vary, so this module can either simulate the knowledge stored of a weak player or a master player. The second component, the performance module, is composed of three sub-modules:

1. The Piece Saliency Detector: given a chess situation, detect for each piece its saliency. This module is similar to the one used in PERCEIVER;
2. The Pattern Recognizer: given a salient piece and its adjacent pieces, form a group of pieces. Then in a second time, this module tries to match this group of pieces with a known chunk (“chunking” is explained in Section 2.3.2, page 14) stored in the learning component (the EPAM net). If this group is present in the EPAM net (**i.e.** this chunk has been learned), then this group is labeled as a known chunk and encoded in short-term memory;
3. The Information Decoder: from a chunk stored in short-term memory, retrieve each piece included in the chunk with their respective position.

MAPP’s Performance

MAPP’s performance in chunk recognition is tied to the quality of its EPAM Net (**e.g.** the content of its long-term memory). To evaluate MAPP, Simon and Barenfeld constructed different EPAM Net and evaluated the recognition ability with each of them. Finally, these results are compared with an expert chessplayer.

Using the published games of different level, two EPAM Nets were built:

1. Net A contains 447 unique chunks (1526 nodes, 447 terminal nodes);
2. Net B contains A’s 447 chunks plus 125 for a total of 572 chunks (1909 nodes, 572 terminal chunks).

Both A and B nets contain two to seven size long chunks. As a systematic sampling of every chunk is labor-intensive, Simon and Gilmarin agree to say that not all chess chunks are covered from those two nets, but this should remain sufficient as a first investigation.

The first comparison is performed using five chess situations at the twentieth move (taken from published game). MAPP succeeded in reproducing from 39% to 43%. Looking at the best performance of MAPP in these situations, the chess-program recognized, on average 4.4 patterns (chunks) per position, and these patterns contained, on average 2.45 pieces.

A second experiment was done on nine tactical positions (taken from Reinfeld's *Win At Chess* book) where the player can gain a tactical advantage if played well. In these sessions, MAPP's performances ranged from 53 to 55%, which is slightly better than the first experiment. The average number of recognized patterns is approximately the same (4.5), however, the average size has increased, rising to 2.79 on average.

The last experiment consists of four chess positions, all taken at the tenth move of grandmaster games. MAPP achieved 73% of correctly placed pieces. The average number of recognized patterns is 5.75, and their average size is 3.9. These performances are explained by the shape of grandmaster's games, which are often "stereotyped" as they tend to use the same pattern during their openings.

Size of the EPAM Net

These presented results raised intriguing questions for Simon and Gilmarin: "*How many patterns, or chunks, should one memorize in Long-Term Memory in order to recognize completely the immense variety of games position offered by chess games? Moreover, how large should be each pattern?*". Simon and Gilmarin did the math. By only considering plausible and reasonable development, they estimated that the number of chessboard positions could be between 10^{10} and 10^{15} . The number of chunks required to represent these positions is also approximate. Using a strong assumption that all chunks are of the same size of pieces, they conclude that to represent 10^{10} requires 280 chunks of 5 pieces each. MAPP was using 447 chunks for net A and 587 for net B. However, we observe during the running process that, on average, the recognized chunks by MAPP were between 2.5 and 3.9 sizes long. Hence one problem of MAPP performance would lie into his EPAM Net, which is poor in large chunks.

To refine their analysis, Simon and Gilmarin took into account that chunks stored into long-term memory are of different sizes and, most of all, unequal in the frequency of appearance. Using the property of the harmonic distribution, the authors computed that to achieve 75% of accuracy for chessboard taken after the 20th first move, 13.500 chunks of different sizes are required. Finally, they summarized the study by positing that it requires at least 100.000 patterns to reach master and grandmasters performances.

Comparison with Human Performance

MAPP showed quite impressive performance in task memory, but does it behave as a human would perform? Simon and Gilmarin (1973) wanted to add convincing evidence that MAPP theory partially simulates human processes. To do so, they asked two players, a class A of intermediate level and a chess master, to take the memory task. They compared the accuracy performance but also the correlation of well-placed pieced and errors between MAPP and the chess master. For accuracy, the chess master was able to reproduce in average 71% of pieces from 14 positions, where MAPP scored 48% but only 41% for the class A player. Among the well-placed pieces from the chess master, MAPP succeeded in placed 55.5% of them, and most of the time, the recognized patterns from MAPP were familiar chunks for the chess master.

Other results are the "within-" and "between-chunk" relations. By analyzing basic relations (*attacking, defending, same color, same type*) that bonds pieces played successively by subjects when reconstructing a board, they showed that frequencies of recognized "within-" and "between-chunk" relations by MAPP are highly similar to chessmasters' ones.

These comparisons between human chess players of different level and the MAPP system shows surprising similarity in memory organization and chunking of information.

Discussion

To summarize, the theory developed around the MAPP system tries to model the human cognitive process engaged in a memory task. Such a task consists of a succession of very distinct steps, and a MAPP's component models each of them. Table 5.1 gives lists all the steps required by a human player to perform a memory task. For each step, the human's ability to use is named, and the corresponding MAPP component is given.

Steps in Memory Task	Human Components	MAPP Components	MAPP Algorithmic Structure
Perceptual phase (de Groot, 1946)	Lateral Geniculate Nucleus (LGN)	PERCEIVER	List of salient pieces Known chess relations between pieces
Salient Pieces Detection	Chess expertise + Long-Term Memory (LTM)	Salient Piece Detector	Function of 3 criteria
Patterns Detection	Chess expertise + LTM	Pattern Discriminator (EPAM Net)	Tree-like structure
Patterns Coding in Short-Term Memory (STM)	Chunking in STM + LTM	Pattern Discriminator (EPAM Net)	Table of fixed length (7) containing names of chunks
Patterns Decoding	Reading back chunks from STM	Reading EPAM Net	Tree-like structure

Table 5.1: Summary of the architecture of MAPP compared with human's abilities to perform a memory task.

As is it the first investigation, one can notice that MAPP does not fully model some steps. For example, the chess expertise of a player is only represented as the content of the EPAM Net in MAPP. Moreover, the Saliency Piece Detector's function is not entirely situation-dependent. This module is based on three criteria, and two among them are related to nearby allied pieces. We know from chess players that a piece can be salient in situations that do not match these two criteria (e.g., a Bishop trapping a King on a corner). Such situation-based criteria may have been learned through years of practice by chess players, and they rely on them to create new patterns in their long-term memory.

In conclusion, by elaborating PERCEIVER and later the MAPP-system, Simon and colleagues described how one could reproduce human cognitive processes engaged in a memory task. The equivalence found in their system's performance and human chess players of different levels suggest that their propositions are relevant and consistent with the underlying human processes.

5.1.3 What Can we Learn from these Systems?

It is interesting to see how Simon *et al.* iteratively processed to elaborate two systems to reproduce human cognitive processes in the 1970s. By analyzing findings from De Groot (1978) and other studies, they developed sub-modules one by one. First, with PERCEIVER (1969), they focused on information extraction with visual features. However, something was missing: an efficient method to encode and recognize the information. To improve this initial system, the authors elaborated on the EPAM (Elementary Perceiver And Memorizer) component to learn relevant patterns. A few years later, Simon *et al.* proposed the MAPP system with more sub-modules: the pattern learner, the salient piece detector, the pattern discriminator, and the chunk decoder.

Is it possible to observe the cognitive processes of chess players similar to the ones proposed by Simon *et al.*? Probably, considering that they were elaborated by observing players' fixations and their verbal protocol. With more modalities, would it be possible to observe more complex cognitive processes?

5.2 A Multimodal Instrument to Observe Human Reasoning

Base on the explored literatures (Chapter 2 and 3), on possible metrics to observe manifestations of reasoning (Chapter 4) and on previous work from Simon *et al*, we designed a recording instrument. It has been made with the following guidelines in mind:

- **Reproducible**;
- **Off-The-Shelf** and **Affordable** sensors;
- **Non-Invasive** (remote) sensors;
- **State-of-the-Art** softwares;
- **Open-Source** softwares.

Modality	Hardware	Freq (Hz)	Software
Body	Kinect 2.0	30	Openpose
Emotion	Webcam	60	1) Noldus FaceReader 2) OpenFace
Gaze	Remote Eye-Tracking Bar	60	1) Tobii 2) Fovio
-	Touch Screen	60	RGBD Sync SDK

Table 5.2: Summary of Hardware and Software used in our Recording Instrument.

Table 5.2 gives a summary of the components (hardwares and softwares) used in our setup. Hardware components are presented in Section 5.2.2 and a description of software used is given in Section 5.2.3.

5.2.1 Domain of Application and Limitations

We want to capture the cognitive processes of chess players engaged in problem-solving in their comfort zone. Ideally, recording two players of the same level, facing each other would be the best environmental setup. Our modalities are emotion, body, and gaze; each of them can be reliably recorded with the use of wearable sensors (electrodes, glasses, motion capture sensors etc.). However, these invasive sensors may affect the players' stress level and take him out of their comfort zone.

To reliably capture information, remote (non-wearable) sensors such as 2D and 3D cameras, and remote eye-tracking must be positioned in front of a player. This is not practical when two players are positioned face-to-face over a chessboard. The use of an interactive touch-sensitive computer screen provides a more easily controllable environment. Also, this arrangement allows us to control the experimental conditions by providing problems with known solutions and levels of difficulty.



Figure 5.3: Designed Experimental Equipment for the CEEGE project, adapted from the *Figurines* project (Portaz et al., 2017). From top to down, sensors are: a Kinect2, a webcam, a touch-screen and a remote eye-tracking (below the screen). All these remote sensors are focusing on the subject engaged in problem-solving presented on the screen.

5.2.2 Hardware

This system drew on lessons learned from an earlier system used to record children during story-telling sessions (Portaz et al., 2017). Our multimodal recording instrument, displayed in Figure 5.3 is composed of several hardware elements that are focusing on a different point of interest of the chess player:

- **Touch-Screen** computer of 23.8 inches;
- **Kinect 2.0** records a 640x480 RGB image with depth information at 30 Hz. Mounted 35cm above the screen and is inclined to capture the upper-*body* part, from the waist to the head included arms, of the chess player.
- **Webcam** records a 1920x1080 RBG image at 60 Hz. Mounted 25cm above the screen, the camera has a frontal view focusing on players' *face*.
- **Eye-Tracking** bar (remote). Mounted either below the screen, horizontally, or vertically on the left of the screen. It captures *eye-gaze* at 60 Hz.
- Two **USB-LED**, adjustable, are used for lighting condition control.

A wooden superstructure is used to rigidly mount the measuring equipment with respect to the screen in order to assure identical sensor placement and orientation for all recordings. This structure has been made using a laser cutter. The Touch-Screen was chosen to provide a gesture-based interaction resembling play with a physical board.

5.2.3 Software

We have used state-of-the-art software for detecting and analyzing human engaged in problem-solving. Except for the eye-tracking solution, we have relied on open-source software so that our experiments can be easily reproduced. All data are recorded synchronously using an enhanced version of the MobileRGB-D software (Vaufreydaz and Nègre, 2014) and analysis are performed offline.

5.2.3.1 Emotion Detection

Currently, only a few free tools are available for detection and recording of facial action unit activations. Some commercial systems exist, but their lack of transparency on the used algorithm and intermediates variables prevent full flexibility in their usability (Baltrušaitis et al., 2018). At the beginning of this work, our team had access to FaceReader software from the company Noldus. This software has been used during our first experiment (see Section 5.4) before moving on to another software OpenFace for the second experiment (see Section 7.4).

Noldus FaceReader

The Noldus¹ FaceReader software is a facial expression recognition software that the entire process of emotion detection from an RGB video. Key paper on FaceReader is Den Uyl and Van Kuilenburg (2005). The used version (7.0), could work both online and offline. Information, presented in real-time during processing, is displayed on a graphical interface, presented in Figure 5.4. Finally, a time bar was available to navigate in the video and the data (available only in offline mode).

In practice, FaceReader software analyses video by first applying a face detector (Viola-Jones) to identify a unique face followed by a detection of 20 Facial Action Units (Ekman and Friesen, 1969). Each action unit is assigned a normalized score between 0 and 1 before being combined in subsets to determine basic emotions. In addition, Valence and Arousal are computed and plotted. Finally, besides providing information about emotional responses, FaceReader software also provided a heart rate estimation (not used in our experiments).

FaceReader was tested on two different datasets: the Radboud Faces Database (Langner, Dotsch, Bijlstra, Wigboldus, Hawk, and Van Knippenberg, 2010) containing 59 different models and the Karolinska Directed Emotional Faces (Goeleven, De Raedt, Leyman, and Verschuere, 2008) which regroups 70 individuals. Both datasets display seven different emotional expressions (plus neutral) from different angles. The FaceReader algorithm correctly classified 90% of the 1197 images from Radboud Face Database and 89% of the Karolinska Dataset (4900 images).

Limits have been observed during our first experiment (discussed in 5.5). Some players' faces were not successfully identified because of small occlusions (**e.g.** hand on chin or cheek). Our hypothesis is that this is the result of the use by Noldus of the outdated Viola-Jones Face detector in Version of FaceReader. FaceReader's pipeline does not provide the flexibility needed to include an alternate Face detection algorithm in the processing pipeline.

OpenFace

OpenFace² (version 2.0) is a framework toolkit that implements many features detection for facial behavior: head pose tracking, eye-gaze, facial landmarks detection and facial Action Units recognition. It was initially developed by Tadas Baltrušaitis in collaboration with CMU MultiComp Lab

¹<https://www.noldus.com/>

²<https://github.com/TadasBaltrušaitis/OpenFace>

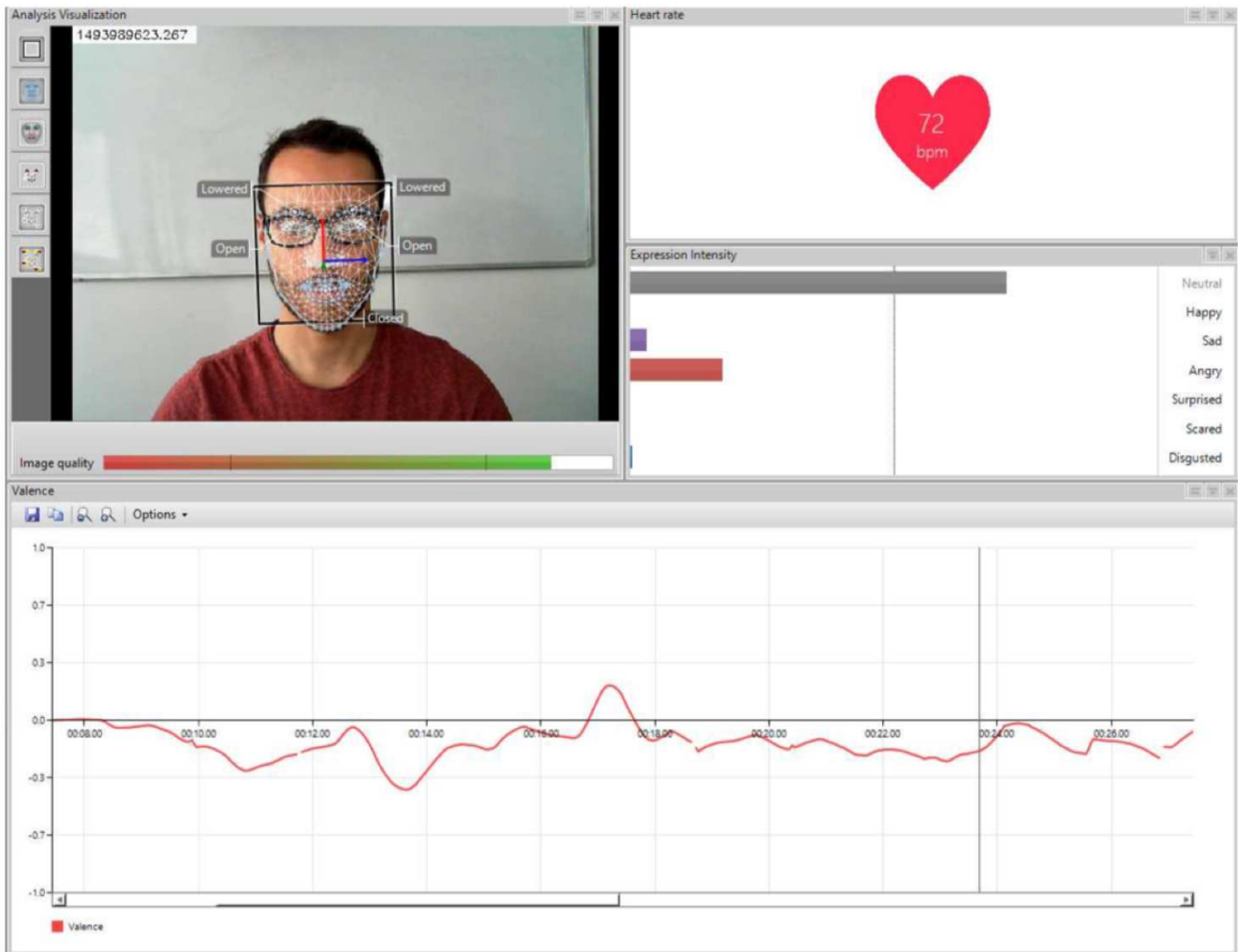


Figure 5.4: FaceReader 7.0 Graphical Interface User.



Figure 5.5: Facial Landmarks Detection with Head Pose Tracking (Left) and Action Units Detection (Right) Performed by OpenFace 2.0 (Source: Baltrusaitis et al., 2018)

led by Prof. Louis-Philippe Morency, the key publication is Baltrusaitis et al. (2018). OpenFace’s source code freely available for research purposes (including training models). Figure 4.14 (see page 49) details OpenFace’s pipeline: landmark detection, head pose and eye gaze estimation, facial action unit recognition (Figure 5.5).

In their publication, Baltrusaitis et al. (2018) present algorithms used for every sub-system coded in the software and compare them with recent approaches. The result of each sub-system outperforms all of the baselines. OpenFace can operate on real-time data video feeds from a webcam, recorded video files, image sequences, and individual images. It is possible to save the outputs of the processed data as CSV files in case of facial landmarks, shape parameters, head pose, action units, and gaze vectors.

5.2.3.2 Gaze Detection

Free eye-tracking software with good performances is rare. Our team had access to products from the Tobii company³, such as eye-tracking glasses and eye-tracking remote bar (version X2). We performed our first experiment using Tobii’s remote eye-tracking before moving on another provider: SeeingMachines, Inc⁴.

Tobii

Remote Eye-Tracker X2-60 from Tobii company is a 184 mm (7.2”) long bar that can track gaze at 60Hz. It must be placed below the screen or on a desk-stand between 30 to 50cm from the user. The software Tobii Pro Studio is used for recording and analysis of eye gaze data. It can adjust the hardware settings, design a test, calibrate, visualize data (HeatMap, ScanPath, Fixations, Bee Swarm), create and manage Area Of Interests, generate Statistics about these AOIs and export data. Figures 5.6 display examples of data Visualization (HeatMap and ScanPath).

The main eye movement measurements used in Tobii Studio are based on two basic events fixations and mouse clicks. Once AOIs are created, different metrics are available to calculated statistics for each participants:

- Time to First Fixation (seconds);
- First Fixation Duration (seconds);
- Total Fixation Duration (seconds);
- Fixation Count;
- Visit Duration (seconds);
- Visit Count.

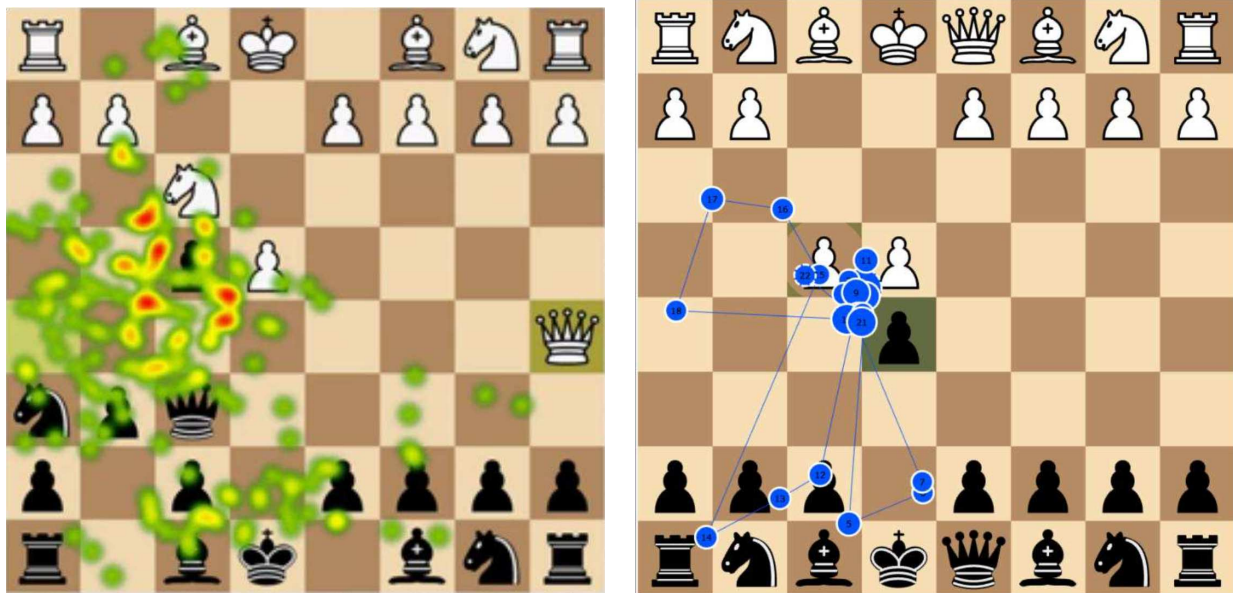
Data can also be exported (csv format) without applying Tobii’s statistics on them. These data can be exported using different pre-processing:

- AOI Gaze events: export gaze events (fixations, saccades) for each AOI.
- Raw gaze coordinates: Export raw X, Y gaze coordinates.
- Media gaze events and coordinates: Export gaze events (fixations, saccades) for each media.

Within our team, people had habits with Tobii’s products. However, during our first experiment with chess players, we have noticed a limitation that is disabling for our case studies. The remote Eye-Tracking bar has to be placed below the screen; this constraint is imposed by the pupil

³<https://www.tobii.com>

⁴<http://www.eyetracking.com/Software/EyeWorks/FOVIO-Simulation-Solution>



(a) Visualization of HeatMap using Tobii Pro Studio.

(b) Visualization of Scan Path using Tobii Pro Studio.

Figure 5.6: Visualization of Tobii Pro Studio (HeatMap and ScanPath)

detection algorithm used by Tobii. Indeed, this algorithm is designed to detect eyes' pupil from below the user (see Figure 4.6b in page 42). This is problematic for our case study. Chess players are interacting with the Touch-Screen computer. Their arms naturally enter in conflict with the Eye-tracking's field of view, covering their eyes from being detected. The bar can not be rotated; neither can it be shifted away too much from the center as it needs to cover the entire screen size. Having no other solution than to use it in a standard way, we have decided to shift the bar eight centimeters to the left. More than eight centimeters causes the eye-tracking from both losing information from the right side of the screen and having trouble to detect the player's right eye. Still, moving it on the left allowed a right-handed player to interact with the screen without disturbing (too much) the proper functioning of the tracker.

Nevertheless, after the completion of the first experiment, we began a search for a more suitable tracker for our test environment. Characteristics of another hardware, Fovio's eye-tracker, designed by Seeing-Machines company, have drawn our attention. Indeed, it can be rotated and does not require to be mounted below the screen.

Seeing Machines

The Eye-Tracker Fovio, provided by *EyeTracking Inc*⁵ and powered by *SeeingMachines*⁶, is a remote bar with a sample rate at 60 Hz. It offers two characteristics that distinguish it from other trackers: 1) the bar can be rotated and mounted vertically or horizontally, and 2) can track simultaneously multiple screens.

EyeTracking Inc proposes a software, named as "EyeWorks", which is a software package developed by researchers for research purposes and designed to work with most existing eye-trackers. This package is composed of the following software:

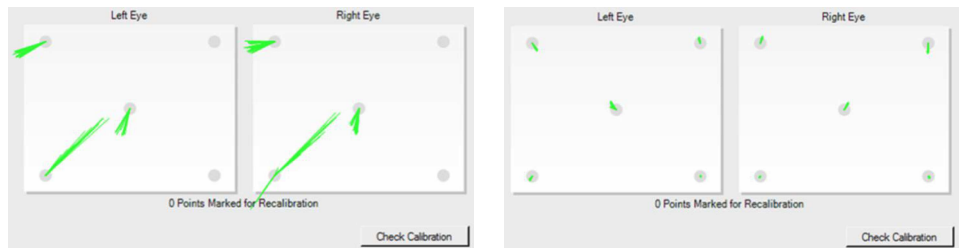
⁵<http://www.eyetracking.com/Software/EyeWorks/FOVIO-Simulation-Solution>
⁶<https://www.seeingmachines.com/>

- **EyeWorks Design:** used to construct a testing script.
- **EyeWorks Record** for multiple uses:
 - Data Collection: Capture eye data, mouse clicks, scrolling, key presses, record audio, video etc.
 - Viewing Options: Real-time viewing of on-screen eye movement data, live streaming for remote viewing etc.
 - Advanced Features: Scene Camera and multi-screen data collection, real-time cognitive workload rendering etc.
- **EyeWorks Analyze:** provides tools for basics eye-tracking analysis (gaze position, fixations, task time), data visualization (GazeSpots, GazeTraces, Bee Swarm Videos etc.).

The main advantage that Fovio brings to our case is its ability to be rotated. We have been able to mount the eye-tracker vertically on the left side of the computer, inclined towards the player. It allows the player to move their hands freely. However, the software “EyeWorks” is, in our opinion, less user-friendly than Tobii Pro Studio.

Calibration

Both of the remote eye-tracking bars used in this document required a pre-phase of calibration for each participant. Most known eye-tracking software relies on the same calibration procedure: 5 to 9 points are displayed on the screen, the user has to fixate the points one by one so that the tracker can carry out the calibration. Figure 5.7 depicted a calibration procedure along with the calibration validation. These images are from Tobii Pro Studio, but EyeWorks rely on the same calibration process.



(a) Bad Calibration Visualization. (b) Good Calibration Visualization.

Figure 5.7: Tobii Calibration Procedure.

5.2.3.3 Body Detection: OpenPose

The chess player’s body is captured using a Kinect 2.0, which is mounted on the top of our recording instrument. This captures two streams of information: RGB and Depth. Microsoft provides, along with the Kinect, a body detection system that detects and annotates multiple bodies on both streams of information. After a few tries, we noticed detection errors when body parts were not visible such as: when the player lays his hands on his knees, under the table, the entire body was poorly recognized, or when the arms of the chair were identified as the player’s arms. Thus, we naturally moved to another software. The state-of-the-art and open source software,



Figure 5.8: OpenPose Body Detection Visualization.

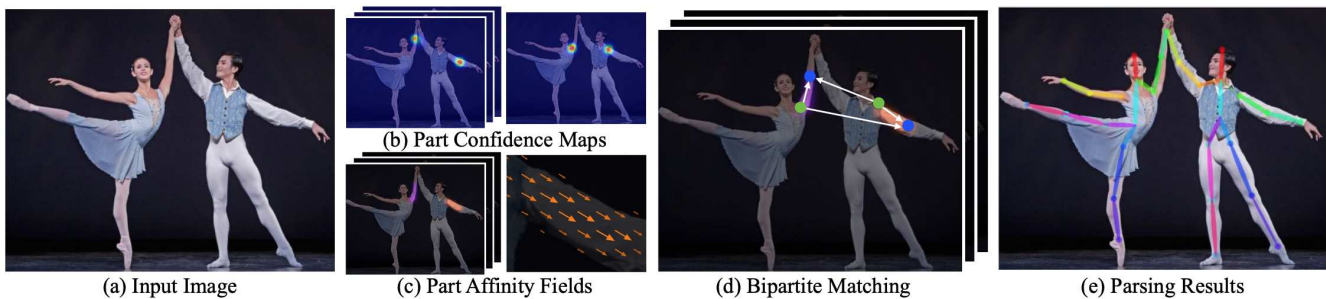


Figure 5.9: OpenPose Overall Pipeline (Source: Cao et al., 2018).

OpenPose⁷, is a real-time multi-person system that is able to: detect human body, hand, facial, and foot keypoints from images or videos. Key publication is Cao et al. (2018). Figure 5.8 presents OpenPose’s pipeline which creates 2D locations of anatomical keypoints (for each person) are created from a 2D color image. To do so, 2D confidence maps of body part locations and 2D vector fields of Part Affinity Fields (which the authors refer to as *PAFs*) are predicted using a feedforward network. PAFs are 2D vector fields that encode location and orientation information across the region of support of the limb. From these confidence maps and PAFs vectors, a greedy inference is used to generate 2D keypoints of body locations for all body present (detected) in the image.

Three different benchmarks have been used to test and evaluate OpenPose against other state-of-the-art methods:

- MPII human multi-person dataset⁸, which contains annotation of 14 body parts for each human body present on images;
- COCO keypoints challenge dataset⁹, which contains annotation of 18 keypoints over 12 body parts alongside 5 facial keypoints, for each human body present on images (Lin, Maire, Belongie, Bourdev, Girshick, Hays, Perona, Ramanan, Dollár, and Zitnick, 2014);
- Subset of 15k images from the COCO dataset with an additional six keypoints annotations for the foot.

⁷<https://github.com/CMU-Perceptual-Computing-Lab/openpose>

⁸MPII human multi-person dataset: <http://human-pose.mpi-inf.mpg.de/>

⁹COCO dataset: <http://cocodataset.org/>

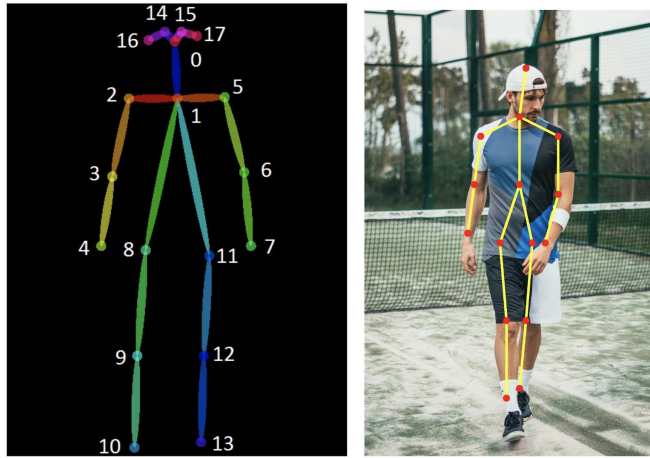


Figure 5.10: Keypoints Detected by OpenPose as Described in COCO Dataset (Source: Lin et al., 2014).

For each dataset, OpenPose has been compared to other state-of-the-art algorithms and achieved high accuracy (if not the highest for some) with the best balance between speed and accuracy (which is one of the main concerns of some algorithm). For our study, the relevant information is body keypoints (see Figure 5.10) of the chess players which are the same used in the well known COCO¹⁰ challenge:

- | | | |
|-------------------|-----------------|----------------|
| 1. Nose | 7. Left Elbow | 13. Left Knee |
| 2. Neck | 8. Left Wrist | 14. Left Ankle |
| 3. Right Shoulder | 9. Right Hip | 15. Right Eye |
| 4. Right Elbow | 10. Right Knee | 16. Left Eye |
| 5. Right Wrist | 11. Right Ankle | 17. Right Ear |
| 6. Left Shoulder | 12. Left Hip | 18. Left Ear |

Keypoints information retrieved by OpenPose can be exported in JSON format. These data are in 2D space (X, Y), so in order to have 3D information, we need to combine information from OpenPose and the Depth (captured by the Kinect 2.0). This step is explained in the Appendix A page 145.

5.2.3.4 Data Synchronization and Visualization

During the study, data were recorded from all sensors (Kinect 2, Webcam, Screen capture, user clicks, Tobii-Bar) using the RGBD Sync SDK¹¹ from the MobileRGBD project (Vaufreydaz and Nègre, 2014). This framework permits to read recorded and further computed data (gaze fixation, emotion detection, body skeleton position, etc.) for synchronous analysis by associating a timestamp with millisecond precision to each recorded frame. The same framework can read, analyze, and display the same way all gathered or computed data. An example is presented in Figure 5.11, where most of the data are depicted.

¹⁰<https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/output.md>

¹¹ <https://github.com/Vaufreyd/RGBDSyncSDK> (last seen 01/2020)

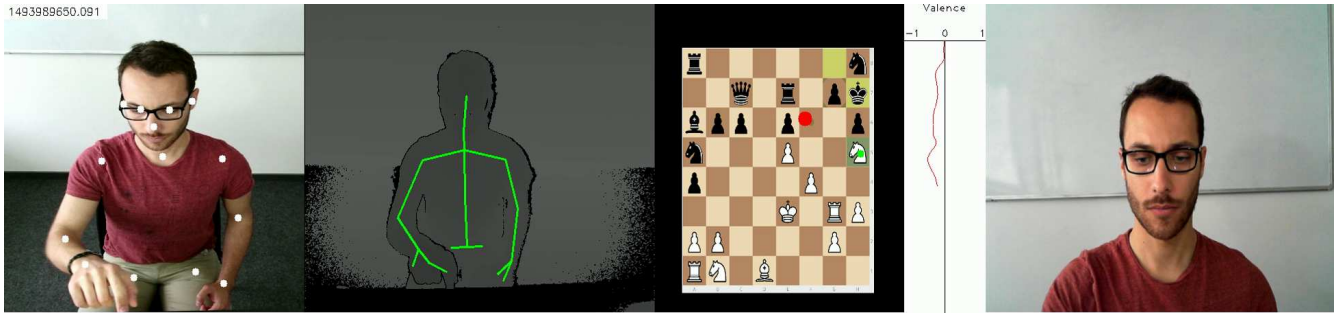


Figure 5.11: Multimodal Visualization of Recorded Data. Left to right: RGB (with body joints) and depth view from Kinect 2 sensors, screen record of chess task (red point is current position of gaze, green point is position of last mouse click), plot of current level of positive emotion expression (valence) and frontal view of face from webcam sensor (Guntz et al., 2018a).

5.2.3.5 Chess Web Platform

The Lichess Web Platform¹² serves for playing and recording games. Lichess webmaster gave their approval to use their platform for this scientific experiment. Most images displaying chessboard in this thesis come from the Lichess Web Platform.

5.3 First Experiment Involving the Local Chess Club

5.3.1 Experimental Hypothesis

Once the recording instrument ready, we elaborated a first experiment to record chess players, like Nemo and Elliott (see Section 2.1), to validate if the setup can reliably capture metrics that reflect reasoning (as discussed in Chapter 4). Our initial research question was:

- Can our experimental set up be used to capture reliable recordings for such study?

If successful, this should allow us to a second research question:

- Can we detect when chess players are challenged beyond their abilities from such measurements, and what are the most relevant features?

Once these first steps validated, we would be able to ask further in depth questions.

5.3.2 Participants and Data Acquisition

Data acquisitions have been made in two sessions in 2017. It has been carried out in the MSH-Alpes¹³ building on the campus of Saint-Martin-d’Hères in March, and 23 participants volunteered to participate. The second session happened during a Chess tournament (14E Open) held in Villard-de-Lans in June 2017, 13 chess players attended the experiment.

Some statistics on the recorded sessions: the average recording time per participant is 13:35 minutes ($MIN = 4:54$, $MAX = 23:54$, $SD = 5:02$) and the average compressed size of gathered data is 56.12 GiB per session.

¹² <https://en.lichess.org/> (last seen 01/2020)

¹³ <https://www.msh-alpes.fr/>

First Session: March 2017

An announcement for the experiment with an invitation to participate was communicated to chess clubs, on the local university campus, and within the greater metropolitan area. We received a positive response from the president, Isabelle Billard, from one of the top metropolitan area chess clubs, and 23 members volunteered to participate in the experiment. Unfortunately, of these initial 23 participants, 9 recordings were not usable due to technical issues. This is discussed in the beginning of Section 5.4. The 14 remaining chess players in our study were 7 experts and 7 intermediates level players (20-45 years, 1 female, age: $M = 31.71$; $SD = 7.57$). Expert players were all active players and with Elo^{14} ratings ranged from 1759 to 2150 ($M = 1950.3$; $SD = 130.0$). Typically, experts are rated between 2000 and 2200 points; masters are between 2200 and 2399; grandmasters are above 2500. For the intermediate players, the Elo ratings ranged from 1399 to 1513 ($M = 1415.3$; $SD = 43.5$) and 6 among them were casual players who were not currently playing in club.

Second Session: June 2017

Although the initial session of data acquisition gave enough interesting results to publish a paper (Guntz, Balzarini, Vaufreydaz, and Crowley, 2017), we knew that gathering more data would allow us to consolidate our results. Being in close contact with the chess club president, she kindly invites us to be present at an upcoming chess tournament taking place in a nearby town. We took this opportunity to communicate about our work and our constant search for volunteers. The project has attracted the attention of many players, and 12 of them accepted to participate in the experiment. Again, technical issues have made us impossible to record 3 players. The 9 remaining players were 2 experts (Elo ratings: $M = 2060.3$; $SD = 50.2$) and 7 intermediates (Elo ratings: $M = 1415.3$; $SD = 43.5$). As they were all also participating in the chess tournament, they were active players (10 – 51 years, 2 female, age: $M=30.0$; $SD=12.9$).

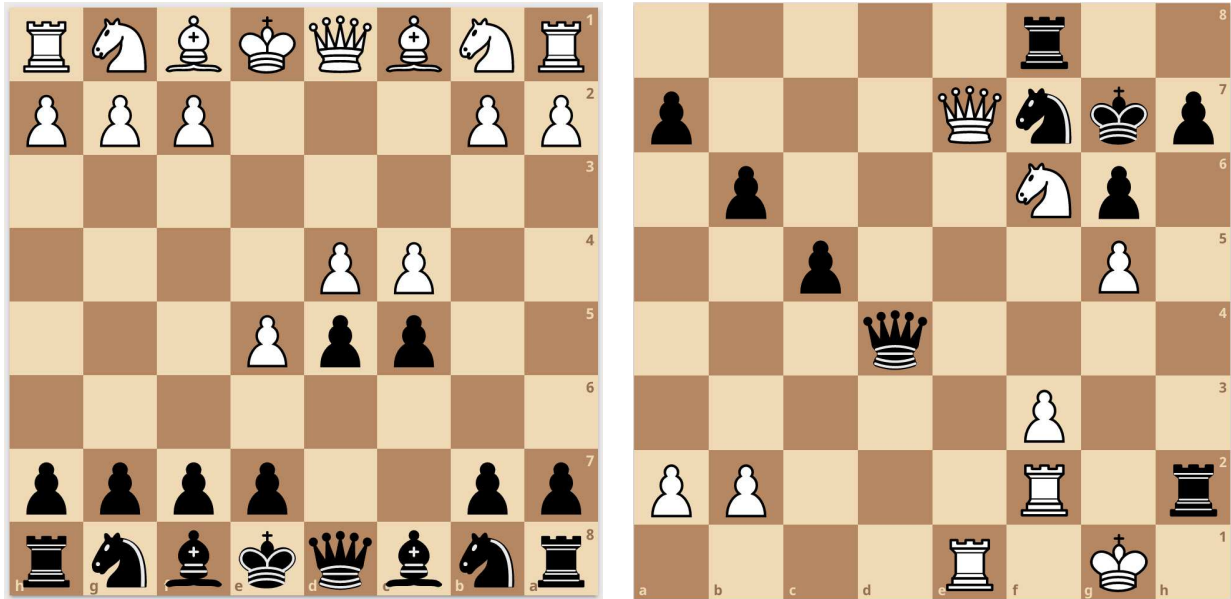
5.3.3 Designing Chess Tasks

The goal of this first experiment was to engage participants in a cognitive process while observing their physiological reactions. To do so, chess puzzles, referred to in this document as *chess tasks*, were elaborated with an increasing level of difficulty. These tasks have been discussed with the collaboration of a coworker from the CITEC team of Bielefeld, Thomas Küchelmann, who is also an active chess player and who contributed to the elaboration of an internal document named as: “*Chess Riddles for CEEGE - Collection of Classified Chess Tasks*”. This document has been a valuable resource for the selection and discussion of chess tasks. Finally, final choices of tasks have been made with the coordination with the president of the local chess club.

Two kinds of tasks were selected: *chess openings tasks*, where only 3 to 5 moves were played from the original state, and *N-Check-Mate tasks*, where 1 to 6 moves were required to checkmate the opponent (and finish the game). Figures 5.12 displayed an example of each task category.

Openings. Skilled players are familiar with most of the chess openings and play them intuitively. Intuitive play does not generally require cognitive engagement for reasoning. A significant challenge is to detect when a player passes from an intuitive reaction to a known opening to challenging situations. Thus, two uncommon openings were selected to this end: a King’s Gambit (3 moves from the initial state) and a Custom Advanced Variation of the Caro-Kann Defense (six moves from the initial state). The goal here is to pull participants out from their comfort zone

¹⁴Elo ratings: see note page 8



(a) Advanced Variation of the Caro-Kann Defense Opening. Black to Play. (b) N-Check-Mate Task with N=3. White to Play.

Figure 5.12: Example of Designed Opening Task (left) and N-Check-Mate Task (right)

as much as possible to evoke emotions and physiological reactions. Openings correspond to task number 1 and 2.

N-Check-Mate. Eleven end game tasks were defined. These are similar to the daily chess puzzles that can be found in magazines or on chess websites. Each of these tasks was designed to checkmate the opponent in a number of predefined moves ranging from 1 to 6. Tasks requesting 1 to 3 moves are viewed as easy tasks, whereas 4 to 6 moves tasks require more chess reasoning abilities, etc. Distribution among the 11 tasks differs according to the number of required moves and thus to their difficulty: 4 tasks with one move, 4 tasks with two and three moves (2 of each) and 3 tasks with four, five and six moves (1 of each). End games were presented to participants in this order of increasing difficulty while alternating the played color (White and Black) between each task.

5.3.4 Experimental Protocol

Participants were tested individually in sessions lasting approximately 45 minutes. Each participant was asked to solve the 13 chess tasks, and their behaviors were observed and recorded. To avoid biased behavior, no information was given about the recording equipment. Nevertheless, it was necessary to reveal the presence of the eye-tracker bar to participants in order to perform a calibration step. After providing informed consent, the Lichess web platform was presented, and participants could play a chess game against a weak opponent (*Stockfish*¹⁵ algorithm level 1: lowest level) to gain familiarity with the computer interface. No recording was made during this first game.

Once familiar and comfortable with the platform, the eye-tracking calibration was performed, in which subjects were instructed to sit between 60 and 80cm from the computer screen and to follow a 9-point calibration grid (see Figures 5.7). Participants were requested to avoid large

¹⁵ *Stockfish* is an open-source game engine used in many chess software, including Lichess. [https://en.wikipedia.org/wiki/Stockfish_\(chess\)](https://en.wikipedia.org/wiki/Stockfish_(chess)) (last seen 01/2020)

head movement in order to ensure good eye-tracking quality. Aside from this distance, no other constraints were instructed to participants.

Each task to solve was individually presented, starting with the openings, followed by the N-Check-Mate tasks ranging in increasing difficulty. Participants were instructed to either play a few moves from the opening or to checkmate the opponent (played by *Stockfish* algorithm level 8: the highest level) in the required number of moves. The number of moves needed for the N-Check-Mate tasks was communicated to the subject. A time frame was imposed for each task, however, it was not announced to the participant; they only knew that they have a couple of minutes to solve the task. This time constraint ranges from 2 minutes for the openings and the easiest N-Check-Mate tasks (1-2 moves) to 5 minutes for the hardest ones (3-4-5-6 moves). An announcement was made when only one minute was remaining to solve the task. If the participant could not solve the task within the time frame, the task was considered as failed, and the participant proceeded to the next task. The experiment is considered finished once all tasks were presented to the participant.

5.3.5 Metrics

Our metrics were selected according to the literature described in Chapter 4. They are listed in Table 5.3 and detailed in the following paragraphs.

Modality	Extracted Data	Computed Features	Computation Software
Body	Body keypoints	Body Agitation Body Volume Self-Touches	Openpose Homemade scripts
Emotion	AUs activation	Emotional state Valence Arousal	Noldus FaceReader Homemade scripts
Gaze	Gaze and Fixation data	Scan Path Fixations Duration Fixations Order	Tobii Pro Studio

Table 5.3: List of Metrics for each modality.

Body Metrics

From what we have learned in Chapter 4, we decided to focus on the following body features:

- **Body Agitation:** how much joints vary along x , y and z axis;
- **Body Volume:** space occupied by the 3D bounding box built around joints;
- **Self-Touching:** collisions between wrist-elbow segments and the head.

2D keypoints are extracted using the OpenPose software. The depth information is retrieved by combining (x,y) position with the depth stream information provided by the Kinect (following the procedure described in Appendix A page 145). Finally, body features are computed from 3D data with homemade python script.

Emotion Metrics

As a first experiment, we wanted to confirm that we were able to efficiently capture the affect of chess players. We decided to focus on the following features:

- **Basic Emotion:** happiness, sadness, anger, fear, disgust and surprise (plus one neutral state).
Each state receives a score between 0 and 1 for his detection. The emotion with the highest score is considered as the current affective state of the user.
- **Valence:** a score between 0 (displeasure) and 1 (pleasure);
- **Arousal:** a score between 0 (calm, quiet) and 1 (excited, aroused).

Eye-Gaze Metrics

Area of Interests (AOIs) have been designed for each designed task. These AOIs were centered on key points of the task which are:

- Pieces that must be moved by the player to solve the tasks;
- Pieces that must be considered to solve the tasks (but not moved);
- All squares that belong to the solution (**i.e.** initial position of the pieces and their destination);
- Empty square(s) that play an important role in the task.

From these AOIs, statistics can be computed. These statistics are our metrics:

- **Scan Path:** Succession of time ordered saccades and fixations from start to the end of the task.
- **Fixation Duration:** Position and duration (in seconds) of each fixation;
- **Fixation Order:** Obtained from Scan Path.

5.3.6 Pilot Experiment: Is the Experiment Protocol Valid?

The described protocol has been tested on coworkers from the lab. The only requirement for this pilot experiment was to know how to play chess; no specific rank or competitive level was asked. The goal of this pilot experiment was to check if all hardware and software were working as expected. A handmade script has been developed to launch all recording software in one step. Five coworkers volunteered and passed the experiment.

Unexpected issues

This pilot experiment raised some unanticipated issues at this time:

1. **Time to pass all chess tasks.**

Some participants used all the available time to solve every task. The total recording for this case almost reaches 45 minutes, which raises the following problem: the size of the collected data.

2. Total size of the data collected.

Videos, images, and raw data gathered at a rate of 30 to 60hz for 45 minutes end up to occupy a significant amount of space. The total size of data for 45 minutes of recording, for a single participant, averages between 350 and 375G. Fortunately, data could be efficiently compressed, but this could only be done once the recording test is completed. Indeed, the computer had not enough resources to handle recording and compressing at the same time.

3. Position of eye-tracking bar.

During the pilot study, the remote eye-tracking was mounted centered below the screen. When players interacted with the Touch-Screen, their arms naturally obstructed the eye-tracker's field of vision. Also, in a moment of uncertainty, players tend to freeze their arms in the air to think again. During this time of hesitation, the eye-tracker's vision was completely obstructed, and no data are recorded.

Solutions

To address these issues, we found solutions:

1. External SSD.

To cope with the limited size space of the recording computer, we started to use an external SSD with 2T of free space. We will describe later in this chapter, that external SSD was not sufficient either, due to another unexpected issue.

2. Eye-tracking shifted to the left.

To avoid obstruction from players' arms, the eye-tracker was slightly shifted to the left (eight cm) from the center of the screen. That way, right-handed players does not disturb the tracker's field of view while playing. However, we will see that a shifted tracker is much sensitive to head movements.

Except for these technical issues, which are mostly related to the hardware, all the features have been successfully extracted and computed. Valuable feedbacks have been gathered from volunteered coworkers, such as adjustments to the explanations given to the participants before starting the experiment.

5.4 Results

Technical issues

Despite our pilot experiment and the detection of several hardware problems, we still lost a significant amount of valuable data. Unfortunately, the precautions taken were insufficient, and the reasons for the loss of data are the same as those we identified: the eye-tracker position and data storage. First, poor tracking results came from several reasons: some participants had difficulties in performing good eye-tracking calibration, or eye-tracking process had been disrupted too many times by participants' movement (**e.g.**, head movements, hand covering the eye-tracking viewing field, etc.) during long recording time (over 40 minutes). This happened despite the left-shifted position of the tracker. Although having been told to avoid making large movements, some players became so caught up in solving problems that they forgot instructions and the nearby recording environment. Secondly, we experimented errors when recording directly on the External SSD. Two hypotheses are possible: either the computer has not enough resources to allocate for both

recording and transferring data on an external disk, or the used USB 3.0 port was defective. Unfortunately, this data transfer issue was random, and when it occurred, we were informed only when the recording was stopped. Due to the nature of our experiment, we could not ask participants to retake the test from the beginning.

5.4.1 Unimodal Analysis

Synchronous data for every feature have been extracted from all sensors. Several tasks, like regression over *Elo* ratings or over the time needed to perform a task, could be addressed using these data. Among them, we chose to analyze a classification problem that can be interpreted by a human: Is it possible, by the use of gaze, body, and facial emotion features, to detect if a chess player is an expert or not? This problem is used as an example to obtain a first validation of our data relevancy. It is correlated with whether a chess player is challenging beyond his abilities or not.

This section presents unimodal and multimodal analysis of extracted features to determine chess expertise of players. Only the data recorded for the 11 N-Check-Mate tasks are considered here, outcomes for participants are presented in Table 5.4.

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task 9	Task 10	Task 11
Difficulty of the task	Easy	Easy	Easy	Easy	Easy	Easy	Medium	Medium	Hard	Hard	Hard
Number of moves required to complete the task	1	1	1	1	2	2	3	3	4	5	6
Number of experts who pass the task (/9)	9	8	8	9	9	9	9	8	8	4	1
Number of intermediates who pass the task (/14)	13	9	12	8	12	13	7	6	3	2	0

Table 5.4: Experts and Intermediates results summary for N-Check-Mate tasks. The number of moves required to complete the task is also given. The higher the number of moves needed, the higher the difficulty.

Eye-Gaze

Two AOIs were defined for each task: one AOI is centered on the very first piece to move in the optimal sequence to achieve the checkmate successfully, and the second one on the destination square where this piece has to be moved. Figure 5.13 shows an example for a N-Check-Task with $N = 1$. For this task, the White Bishop in E2 has to be moved on square B5 in order to checkmate the opponent’s Black King and win the game (solving the puzzle). To take into account that players may use their parafoveal vision to extract information, AOIs are twice larger than a regular chess square.

Fixations information of every task is gathered for all participants, and results are presented in Figure 5.14. As can be seen in this figure, experts have longer and more fixations than intermediates on relevant pieces, which is the opposite of what the literature stated (Reingold and Charness, 2005, see Section 4.1.4 page 45 for more information). This difference is mainly due to the chosen AOIs. They are focused on the first piece to be moved, and its destination, whereas Reingold and Charness considered the entire board to collect players’ fixations. Another argument can be put forward to explain this difference with the literature.

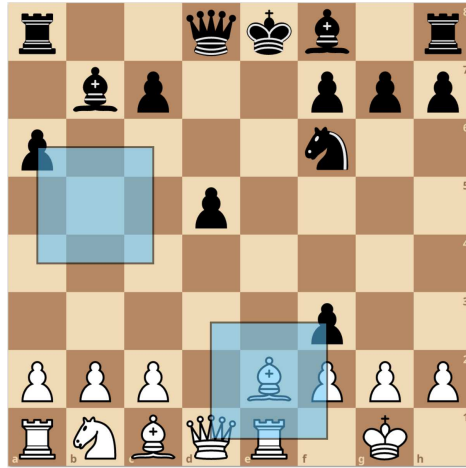


Figure 5.13: AOIs, designed for a N-Check-Task ($N=1$), are Centered on the First Piece to be Played (White Bishop in E2) and its Destination (B5). AOIs are larger than a Square to Capture Fixations that would use Parafoveal Vision.

For this, we need to refer to the four stages of the thought process for chess players proposed by De Groot (1978):

1. The orientation phase: participants scan the board to grasp information about pieces' organization;
2. The exploration phase: participants consider variations (moves) from the current configuration;
3. The investigation phase: participants analyze in-depth the two most probable candidates from phase 2;
4. The proof phase: participants confirm the validity of their choice.

In our case, intermediates have fewer fixation times in proportion on relevant pieces; one hypothesis could be that these participants spend more time during the first phase or do not consider the correct pieces fast enough during the investigation phase. On the other side, experts tend to spend more time on relevant pieces. This statement conforms with a valuable finding from De Groot, who stated that experts are more efficient and faster to identify key pieces.

In our results, the difference in fixation duration between experts and intermediates is statistically significant ($p < 0.05$) and could be used as a discriminant feature. This is explained by experts' skill encoding capacity that enables them to quickly focus their attention on relevant pieces by a better pattern matching ability. More valuable information could have been extracted from players' gaze, especially considering the literature on the subject (Charness et al., 2001; Reingold and Charness, 2005).

Emotions

The increasing difficulty in the non-interrupting tasks has caused our participants to express many observable emotions across the experiment. Emotions in a long-task experiment are expressed as peaks in the two-dimensional space (valence, arousal). Thus, standard statistics tend to shrink toward zero as the record becomes longer. Other approaches should be considered to visualize emotional expressions. One possibility is to consider the number of changes of emotions having

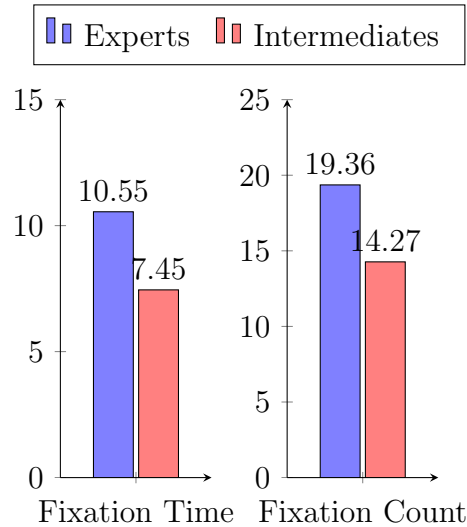


Figure 5.14: Eye-gaze Histograms. Left: Percentage of fixation (in seconds) for all tasks. Right: average over all tasks of the number of fixations.

the highest intensity (*i.e.* the current detected facial emotion). As emotion intensities are based on facial unit detection, changes in the main emotion denote underlying changes in facial expression. The result metric is shown on the graph presented in Figure 5.15a.

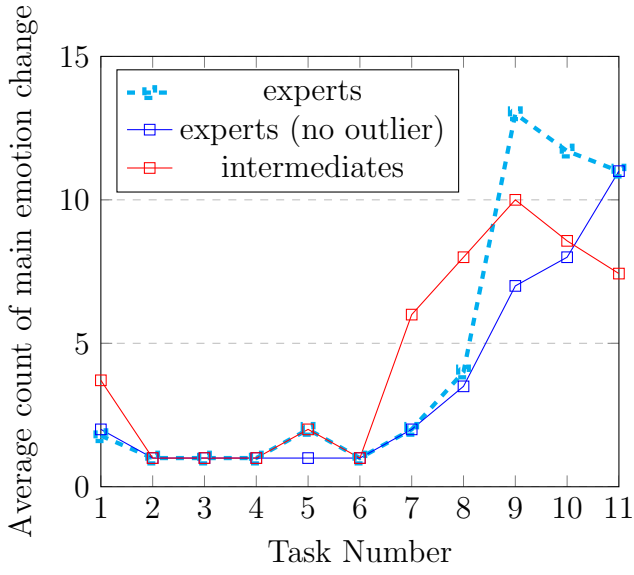
It appears that expression of emotions increases with the difficulty of the problem to solve. For both group levels, experts and intermediates, the number of observable emotions increase when participants started to be challenged with difficult tasks. As can be observed, all lines remain close to 0 for tasks requiring less than 3 moves to checkmate the opponent (below task number 6). From task 7 (3 moves), participants started to be challenged. Among the expert group, one participant presented surprising results with a high number of emotion variations. This participant could be interpreted as an outlier as more than twice the average of variations has been detected for him. To illustrate this, two lines have been plotted, the cyan one represents all experts, whereas the blue one omits this potential outlier. To support this outlier hypothesis, the resulting blue line (along with the red line for intermediates participants) is similar to a result we showed during the same experiment environment with 14 participants (see figure 5.15b).

Another interesting aspect of that plot is the final decrease for both groups after task 9; this could be interpreted as a sort of resignation when players knew that tasks beyond their skills and could not be resolved. Indeed, as shown in Table 5.4 (page 78), only 6 participants (2 intermediates and 4 experts) passed the 10th task, and only one expert solved the last task. These results suggest that situation understanding and expertise knowledge can be inferred from variations of facial emotions. More discussion about emotion is given in Section 5.5.

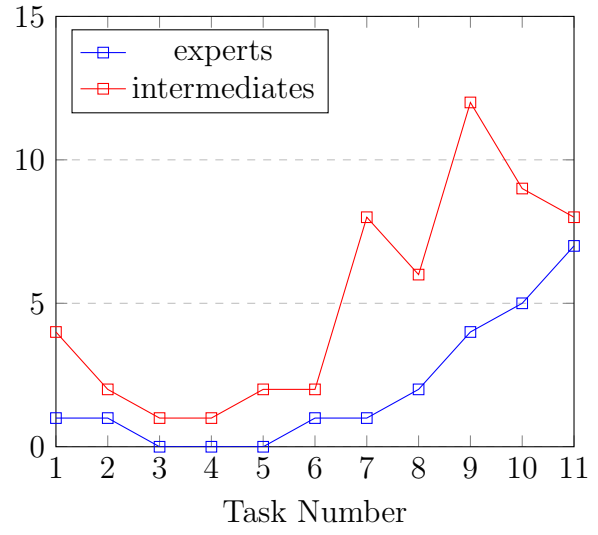
Body Posture

We can observe how participants' body reacts to the increasing difficulty of tasks with Figure 5.16 that presents statistics about self-touching.

Similarly to the results on emotions, a variation in participants' behavior is observed from the first task requiring more than 2 moves (task 7) to be completed. This change is observed as an increase of self-touches, especially for intermediates, whereas experts' line shape looks more like the beginning of an exponential curve. Thus, a first observation is that the number of self-



(a) Current study analysis on 23 participants.



(b) Previous study realized on 14 participants.

Figure 5.15: Average Count of Variation of Main Detected Facial Emotion in Regard to the Task (1-11). Tasks are Ranging in an Increasing Difficulty Order.

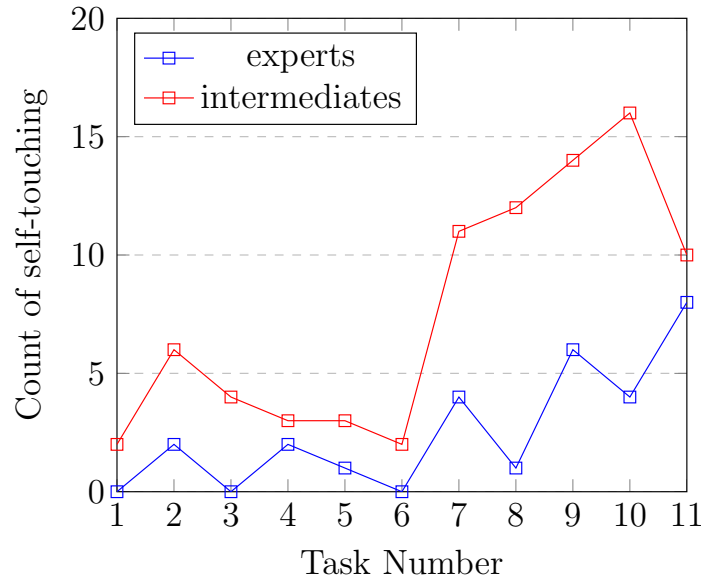


Figure 5.16: Count of Self-Touching in Regard to the Task (1-11). Tasks are Ranging in an Increasing Difficulty Order.

touches increases as tasks get harder, and it reveals that this is a relevant feature to consider. These results can be interpreted by referring to the literature stating that these bodily behaviors occur when subjects are subjected to stress (Harrigan, 1985) and cognitive overload (Troisi, 2002). Furthermore, the difference in the number of self-touches between experts and intermediates is statistically significant ($p < 0.05$) and could be used, as well as other related features on body (like agitation), to perform classification.

5.4.2 Statistical Classification

To demonstrate the potential benefit of a multimodal approach, a supervised machine learning algorithm has been used to quantify the accuracy of different modalities for classification.

Support Vector Machines (SVM) have been built for each modality and for each possible combination of modalities. For this purpose, we computed statistical analysis with respect to the nature of our data: discrete (count, duration) or continuous (mean, variance, standard deviation) over our metrics. Giving us a total of 3 features for the gaze modality, 30 for the emotion modality, and 14 for body modality (see Table 5.5). Then, we constructed input samples as the following: one input sample would be the instantiation of one participant for one particular task, given a total number of $23 \times 11 = 253$ input samples.

Modality	Metrics	Data type	Statistical transformation	Number of Features
Gaze	Fixation	Discrete	Duration - Count	2
	Visit	Discrete	Count	1
Emotion	7 Basics Emotions	Continuous	Mean - Var - Std	21
	Valence	Continuous	Mean - Var - Std	3
	Arousal	Continuous	Mean - Var - Std	3
	Heart Rate	Continuous	Mean - Var - Std	3
Body	Agitation (X, Y, Z)	Continuous	Mean - Var - Std	9
	Volume	Continuous	Mean - Var - Std	3
	Self-Touch	Discrete	Duration - Count	2

Table 5.5: Metrics Overview

A 10-fold stratified cross-validation procedure has been used on every SVM to compute their accuracy. An exhausting grid search over a different set of hyperparameters is used to explore and find the best parameters for each SVM to consider. Two grids were explored, one *RBF* kernel and one *Linear* kernel. The C parameter for both kernels took values in [1, 10, 50, 75, 100, 1000] and the γ for *RBF* kernel ranged in [0.1, 0.01, 0.001, 0.0001, 0.00001, 0.000001]. To compute accuracy for one SVM, each stratification is randomly shuffled before splitting into batches and before computing accuracy. This process is repeated 100 times, and the final accuracy for the SVM is the average of all randomly shuffled stratification. Accuracy results overall features are depicted in the first line of table 5.6.

A first observation is that emotion and body modalities reached good accuracies, 86% and 90% respectively, confirming that these features capture relevant information. Gaze modality performed less than expected, only 62%; this could be explained by the way the metrics have been

Modalities	G	B	E	G + B	G + E	B + E	G + B + E
Number of Features	3	14	30	17	33	44	47
Accuracy Score	0.62	0.90	0.86	0.81	0.86	0.83	0.83
Standard Deviation	0.24	0.09	0.17	0.14	0.16	0.12	0.19

Table 5.6: Best Accuracy Scores from Cross-Validation for SVMs (10-Fold Cross Validation). The number of sample N is the number of participants (23) times the number of N-Check-Mate tasks (11). Columns are the modality subset chosen to train the SVM (G: Gaze, B: Body, E: Emotion).

chosen. Indeed, we decided first to analyze fixations for only 2 AOIs without considering scan path order.

Secondly, modality fusion varied between 81% and 86% for any possible combination, which is less than a unimodal approach. To perform expertise classification, one might be tempted to choose only body or emotions features. However, a system relying on a unique modality may face unreliable or noisy data in real-life conditions. For example, emotion detection can not be performed if there are occlusions in front of the subject head. Being able to rely on multi modalities to build a robust system is essential.

5.4.3 Features Selection

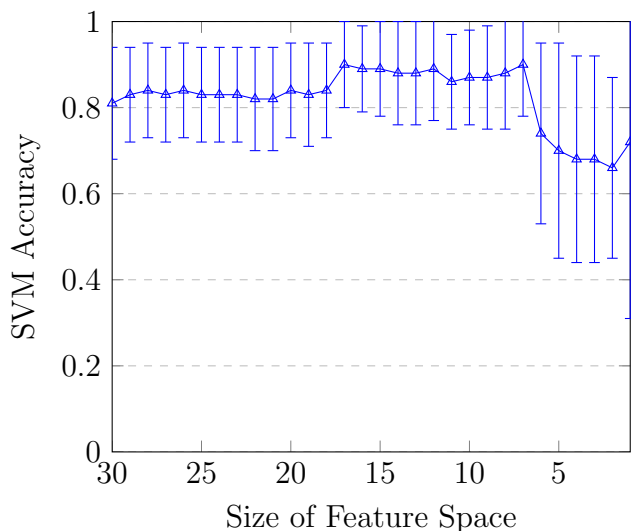
We investigated which feature is more robust for our system. As several features characterize each modality and our system takes into account 47 features, we propose to evaluate the most relevant for the expertise classification task.

For this purpose, among existing methods in the literature, we selected two different techniques, the first one is the Minimum Redundancy Maximum Relevance (mRMR) technique (Peng, Long, and Ding, 2005) for the expressiveness of the selected features we experienced formerly (Vaufreydaz, Johal, and Combe, 2015) and the second is the Fisher Feature Selection (Duda, Hart, Stork, et al., 1973; Li, Cheng, Wang, Morstatter, Robert, Tang, and Liu, 2016). These algorithms allow a dimensionality reduction of the feature space according to the maximal statistical dependency criterion based on mutual information. Both techniques search to maximize features relevance and effectiveness; however, they differ in their redundancy policy. Indeed, mRMR tends to reduce redundancy as much as possible even if it would impact the overall effectiveness; on the other hand, Fisher may keep redundancy if it can improve effectiveness. We propose here to use both techniques separately on our data and compare results.

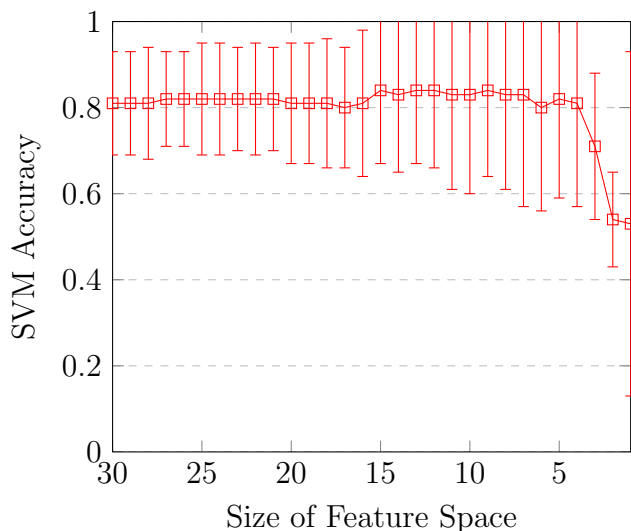
Using these techniques, we expect to reduce the feature space drastically to improve the data fusion performed by the SVM classifier. Both mRMR and Fisher techniques have been used on all our features, and the resulting ranking is listed in table 5.7 and 5.8. Key information here is that among the top-rated selected features, all three modalities are present. It confirmed that all modalities bring relevant information. Fisher’s selection instead prefer to keep body and emotion information as its maximize effectiveness; this has been observed with SVM accuracy (see table 5.6), even if some features are redundant (like agitation in x, y, z) where mRMR preferred to keep only one of them. It is interesting to observe that both techniques agreed to select top features for each modality. The variance about agitation and volume combined with the number of self-touches gives us significant information about body behavior. Variance, standard deviation, and mean about valence, heart rate and - negatives - basics emotions (disgusted, scared and angry) capture emotional reactions the participants when being challenged. Finally, as specified in section 5.4.1, fixation duration is relevant to classify a participant as an expert or intermediate.

With the aim of benefit from both feature selections, the impact on accuracy is showed in Figure 5.17. The highest accuracy score for mRMR selection is 84% using its top 9 ranked features. Using Fisher features selection, accuracy can reach up to 90% with its top 7 ranked features. One interesting observation from these plots is the two different paces for accuracies to reach their maximum. Only 4 non-redundant features combined from mRMR ranking are enough to reach more than 80% of accuracy where the slow but steady curve from Fisher needs to combine 7 features (and go even way higher with 90%). 8 firsts features are enough to reach up to 77%, which is close to the one achieved earlier (82%) by using all features.

These results show that features from different modalities are not equally relevant. Figure 5.17 illustrates that after certain amount of features, performance stagnates or even decreases slightly. This phenomenon starts after 15 and 17 features for mRMR and Fisher techniques, respectively. This can have 2 explanations. The first one is that the number of input data is not high enough to take advantage of these features. The second explanation is that these last features do not actually provide new information for classification. This hypothesis is consistent with the ranking results from both mRMR and Fisher techniques, which classified them last and therefore showed that they contain a large amount of redundant information already included in top features.



(a) Fisher ranking features.



(b) mRMR ranking features

Figure 5.17: Mean with standard Deviation of the Accuracy Scores from 10-Fold Cross-Validation for SVMs with Decreasing Number of Features. Features order is given by Fisher (see Table 5.8) and mRMR (see Table 5.7) algorithms.

mRMR Ranking Order	Feature	Modality	Description
1	Y_Agitation_var	Body	Variation of agitation on Y axis
2	Disgusted_std	Emotion	Standard Deviation of the detected basic emotion: Disgusted
3	Fixation_Duration	Gaze	Average Fixation Duration on AOI
4	Valence_mean	Emotion	Mean of the computed Valence
5	Volume_var	Body	Variance of the body volume
6	HeartRate_std	Emotion	Standard Deviation of Heart Rate
7	Angry_var	Emotion	Variance of the detected basic emotion: Angry
8	SelfTouches_Count	Body	Average number of self-touches
9	Scared_var	Emotion	Variance of the detected basic emotion: Scared
10	Angry_mean	Emotion	Mean of the detected basic emotion: Angry
11	Fixation_Count	Gaze	Average Number of Fixation on AOI
12	X_Agitation_std	Body	Standard Deviation of agitation on X axis
13	Happy_mean	Emotion	Mean of the detected basic emotion: Happy
14	Disgusted_var	Emotion	Variation of the detected basic emotion: Disgusted
15	Volume_std	Body	Standard Deviation of the body volume
16	HeartRate_mean	Emotion	Mean of Heart Rate
17	Sad_std	Emotion	Standard Deviation of the detected basic emotion: Sad
18	Arousal_mean	Emotion	Mean of the computed arousal
19	SelfTouches_Duration	Body	Average duration of self-touches
20	Neutral_var	Emotion	Variation of the detected basic emotion: Neutral

Table 5.7: Ordered List of the 20 Most Relevant Features generated by mRMR algorithm

Fisher Ranking Order	Feature	Modality	Description
1	Valence_mean	Emotion	Mean of the computed Valence
2	Y_Agitation_var	Body	Variation of agitation on Y axis
3	Z_Agitation_var	Body	Variation of agitation on Z axis
4	Y_Agitation_std	Body	Standard Deviation of agitation on Y axis
5	X_Agitation_var	Body	Variation of agitation on X axis
6	Angry_mean	Emotion	Mean of the detected basic emotion: Angry
7	Z_Agitation_std	Body	Standard Deviation of agitation on Z axis
8	X_Agitation_std	Body	Standard Deviation of agitation on X axis
9	Volume_mean	Body	Mean of the body volume
10	HeartRate_mean	Emotion	Mean of Heart Rate
11	Disgusted_std	Emotion	Standard Deviation of the detected basic emotion: Disgusted
12	Angry_var	Emotion	Variance of the detected basic emotion: Angry
13	Sad_mean	Emotion	Mean of the detected basic emotion: Sad
14	Fixation_Duration	Gaze	Average Fixation Duration on AOI
15	X_Agitation_mean	Body	Mean of agitation on X axis
16	Y_Agitation_mean	Body	Mean of agitation on Y axis
17	Z_Agitation_mean	Body	Mean of agitation on Z axis
18	Disgusted_var	Emotion	Variance of the detected basic emotion: Disgusted
19	Volume_std	Body	Standard Deviation of the body volume
20	SelfTouches_Count	Body	Average number of self-touches

Table 5.8: Ordered List of the 20 Most Relevant Features Generated by Fisher Algorithm.

5.5 General Discussions

This research and results show interesting results on unimodal features used to distinguish expert and intermediate chess players. Due to the size of our dataset, generalizing these results was not possible at that moment.

Our results showed that negative emotions carry significant information about the participants' situation understanding. However, the semantic used by the FaceReader software (*i.e.*, the 7 basics emotions) should be carefully interpreted. This study does not intend to demonstrate that humans engaged in problem-solving would always express the same basic emotions but would instead show a variation in facial action unit (AU) activations. Thus, even if the ranking Tables 5.7 and 5.8 reveals that variations in negative emotions are relevant, one should understand that action units' activation (involved in these emotions) are varying significantly.

Both feature selections from mRMR and Fisher algorithms give interesting results and improve data fusion alongside with classification accuracy. By only considering relevant features and removing redundancy, mRMR feature selection uses its 4 top features to reach 81% of accuracy. Fisher selection authorizes redundancy if it can increase effectiveness, using 7 features, Fisher selection performs 90%.

One should question the importance of redundancy in a multimodal study. Real-life recordings may experiment with missing or noisy data on one or several modalities. Having redundancy between modality may allow the system to still operating even if some modality is facing unreliable data. Fisher selection shows that redundancy requires more features than non-redundancy selection but can improve accuracy efficiently. On the other hand, if a light system can not process too much data, one can choose only relevant and non-redundant features to guarantee maximum accuracy with the lowest set of features.

The conditions of the chess tasks should also draw attention. In the experimental configuration, chess players were facing a chess algorithm engine in tasks where they knew the existence of a winning sequence of moves. Moreover, players were seated in front of a touched screen and were asked to interact with it for playing. Some clues, like body agitation or body volume, may provide different information in another experimental environment. Also, participants may not be as engaged as they would have been in a real chess tournament facing a human opponent using an actual chessboard. In these particular situations, involving stakes for players, the physiological reactions and emotional expressions are more interesting to observe.

Nevertheless, these experiments reveal that valuable information can be observed from human attention and emotions to determine understanding, awareness, and affective response to chess solving problems. The feature selection gives good insight on where we should put our attention to grasp the most relevant information from every modality. Another underlying result is the validation of our setup in monitoring chess players.

5.5.1 Conclusion

This first experiment presents results with the capture and interpretation of multimodal signals of 23 chess players engaged in solving 13 challenging chess tasks. Reliable recording signals have been made with our experimental equipment composed of a Kinect capturing body posture, a webcam recording variation of facial action units, and finally, an eye-tracker analyzing gaze trajectories. These records contain relevant information about situation awareness of humans engaged in solving 13 challenging chess tasks. In addition to validating our initial question that our set up can capture reliable recordings for such study, this equipment is based on off-the-shelf commercially available components as well as open-source programs and thus can be easily replicated. As well as providing

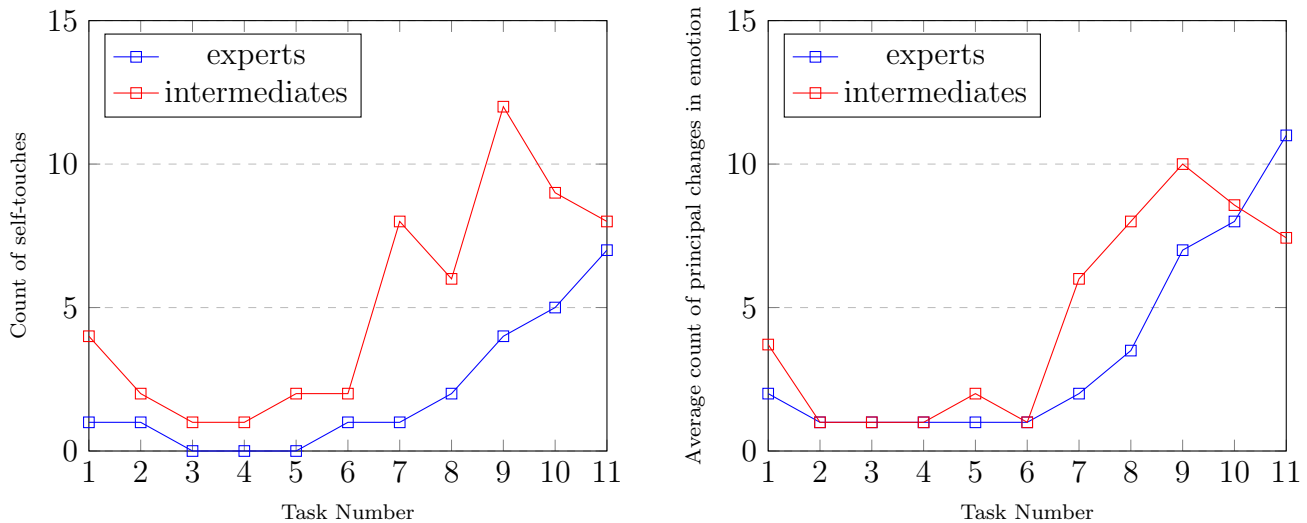


Figure 5.18: Self-Touches (left) and Average Count of Number of Changes in Emotion State (right) for Intermediate and Experts Over the 11 Tasks.

a tool for studies of participants engaged in problem-solving, this equipment can provide a general tool that can be used to study the effectiveness of affective agents in engaging users and evoking emotions.

Afterward, unimodal and multimodal analysis have been performed on these signals to validate our second research question. Fixation durations from eye-gaze reveal the encoding skills of participants according to their expertise level. Self-touches and body posture agitation can be interpreted as a stress reaction when participants are challenged. Finally, variation in emotional responses of participants is correlated with task difficulty. Each of these modalities carries enough information to answer our second research question about detecting when participants are challenged beyond their abilities. Moreover, Support Vector Machine classifiers trained with 10 cross-fold validation revealed that combining such modalities could also give good performances: 90% with only 7 features. This would allow us to build a multimodal system that is more reliable and more stable than systems relying on only one sensor.

5.5.2 Drawback

Figure 5.18 shows the number of self-touches and changes of emotion for intermediate and expert players over our increasingly challenging problem set. Our initial hypothesis was that subjects would exhibit sustained displays of emotions ranging from pleasure to frustration as the difficulty of the problems increased. We were surprised to observe that this was not the case. Rather, both self-touching and rate of change in emotion state evolved from a neutral emotion during reactive play to a period of frequent touching and rapid changes in emotion as the problems became more and more challenging.

Figure 5.18 illustrates that the rate of changes in the emotional state increases with difficulty for both intermediates and experts, with significantly higher numbers for intermediate players. The correlation with the rise in self-touching confirms that subjects were increasingly challenged. We conclude that frustration for intermediate players rose rapidly for tasks 7, 8, 9, and 10, and then dropped, as subjects seemed to abandon efforts to solve task 11. For experts, self-touching and changes in emotion gradually increased for problems 7 through 11, indicating that experts experienced only minor discomfort for these problems.

To interpret these results, we proposed a model (see next Chapter 6). In order to develop our model, we conducted a second experiment in which players were asked to explain their reasoning. The objectives were to determine if eye-gaze, valence, arousal, and frustration could be correlated with the four phases of reasoning proposed by De Groot, and to construct an ontology for chess concepts (chunks and relations) used by players.

Chapter 6

Modeling Awareness from Observation of Eye-Gaze and Emotions

Awareness is the ability to know and perceive memories of events directly. More broadly, it is the state of being conscious of something (definition inspired by Wikipedia contributors, 2020a). When a human is aware of something, knowledge, and memories of that thing are directly accessible. Awareness is often synonymous with consciousness, although the term consciousness can be even more controversial to define.

This chapter presents in Section 6.1 a model that represents knowledge using the Frames theory from Minsky (1974). Using this formal representation, we can elaborate a model of Human Memory that manipulates concepts and knowledge. We finally present a probabilistic approach to evaluate awareness of chunks using evidence of observations (Section 6.2). This model allows us to interpret the results obtained during our first experiment and to develop a second to explore the role of emotion in reasoning.

6.1 A Formal Representation of Chess Concepts and Expertise

As discussed in chapter 2, human abilities for reasoning, and problem-solving are strongly constrained by limits to working memory. A limited number of working memory elements are used to associate perceptions with different forms of Short-Term and Long-Term memories. Minsky provides a formal representation for this process in order to provide a formal, computationally feasible model for reasoning and problem-solving (Minsky, 1974).

6.1.1 Frames: A Formal Representation for Concepts

In 1974, Minsky proposed a theory to unify Artificial Intelligence and Psychology concepts to better model human reasoning. This theory, referred to as the “*Frame theory*”, is used to express formal representations of concepts and is defined as follows:

“When one encounters a new situation (or makes a substantial change in one’s view of the present problem) one selects from memory a structure called a Frame. This is a remembered framework to be adapted to fit reality by changing details as necessary.”

Minsky (1974)

This adaptation, from remembered representation to the perceived world, precisely reflects the general idea of the Situation Modeling Theory previously explained (in Section 2.4.1). Frames are data-structures used to describe a specific situation. They can represent an entity, a relation between frames, or a chunk (*i.e.* composition of sub-frames). A frame can be divided into two levels of information:

- Top-level information: properties and procedures that are always true. This information is internal and immutable about the described situation. Altering one of them necessarily creates a new frame.
- Low-level information: specific data filled during instantiation. These values may vary and may be a reference (pointer) to another frame.

In addition, a frame is labeled with a unique name that can be used as a reference. The recognition of a known concept instantiates a frame representing this particular concept. Top-level information about this concept is retrieved, and low-level one is filled according to the perceived situation. An analogy to Minsky’s frame is object instantiation from *classes* in object-oriented programming (or OOP). In these programming languages, the structure of an object is defined by a class, which is a definition or signature; this is shared among a particular type of object. Upon creation of an object, the properties are set according to the class’s definition. The object is said to be an instantiation of the class.

Let us consider an example: the abstract representation of a cat and two instantiations using frames are given in 6.1. In this thesis, top- and low-level information is represented in bold and italic, respectively.

<pre>(Entity Cat (Cat-ID) (Class (Mammalia)) (Order (Carnivora)) (<i>Gender</i> (one-of (Male, Female))) (<i>Breed</i> (one-of (European Shorthair, Ocicat, Cyprus, etc.))))</pre>	<pre>(Entity Cat (Kovu) (Class (Mammalia)) (Order (Carnivora)) (<i>Gender</i> (Male)) (<i>Breed</i> (Cyprus))) (Entity Cat (Kenya) (Class (Mammalia)) (Order (Carnivora)) (<i>Gender</i> (Female)) (<i>Breed</i> (European Shorthair)))</pre>
--	--

Listing 6.1: Toy entity example: cats

A cat is identified by an ID, has two immutable properties (class and order), and two variable data (gender and breed). The keyword *one-of* specified that upon instantiation, only one value is kept to describe the concept. Two cats are instantiated named *Kovu* and *Kenya*, they both inherited data from the abstract data-structured. *Kovu* is a mammalian and is a carnivora, as all cats, and more precisely, it is a male of the Cyprus breed. About *Kenya*, she is a female and European Shorthair cat.

Now let us have a closer look at relations and sub-frames with the second example 6.2. Here, an abstract representation of a relation, named as “cat sibling”, is presented. Two low-level data are described, the first and the second sibling, which must be filled with a pointer (represent with the symbol “*”) to an object from the specific class Cat. One instantiation of this relation is given and is called “*CatCat*”. Variable data of such instance is set to the existing frame Cat-Kovu and Cat-Kenya, created in the previous example.

<pre> (Relation Cat_Siblings (Cat_Siblings- (<i>Sibling_1</i> (Entity Cat *))) (<i>Sibling_2</i> (Entity Cat *))) </pre>	<pre> (Relation Cat_Siblings (CatCat) (<i>Sibling_1</i> (Kovu *))) (<i>Sibling_2</i> (Kenya *))) </pre>
---	--

Listing 6.2: Toy relation example: sibling cats

From the proposed Frame definition and these two examples, we now build more complex data-structured to both cognitive and chess concepts.

6.1.2 A Representation for Chess Concepts

In 1973, Chase and Simon reproduced and extended researches done by De Groot (Chase and Simon, 1973a). By interviewing three players of different levels (master, advanced and beginner), they stated that the ability to perceive information on the board rapidly and to encode it in short-term memory efficiently is highly related to expertise. It was clear to the authors that players learn relations between pieces in a meaningful context of a standard game. The positioning of pieces only makes sense to players when they are in relation to other pieces. It is these interactions between pieces (or should we call them “chunks?”) that are recognized by the players when presented in a new situation. Besides, Chase and Simon found that master players could recall chunks of average size varying between 1.2 and 3.8, whereas the advanced and beginner player could only retain between 1.0 to 2.5. Five chunks of relations successively placed pieces were listed by the authors:

- Threatening,
- Protecting,
- Proximity,
- Same color,
- Common type (e.g. three Pawns or two Bishops).

Interestingly, two kinds of chunks have been revealed. The first is based on a direct relation between two adjacent pieces and is present in chunks of all players. The second one involves a complex combination of several pieces converging into attacking one or more key pieces of the opponent. This chunk structure is more abstract and less direct than the first one, and only master players could see them under time constraint pressure. On their side, Simon and Gilmartin (1973) estimated that expertise in chess may require between 10,000 and 100,000 chunks in Long-Term memory (discuss in Section 5.2). This rich vocabulary enables players to determine good moves with a moderate search of the game tree.

Let us see how we can apply Minsky’s Frames to chunks learned from Chase and Simon. The basic entities in chess are the individual pieces with properties that can include the type of piece, color (White or Black), and board position. This can be defined as a frame using the following schema:

```

(Entity ChessPiece (ChessPiece-ID)
  (Kind (one-of (king, Queen, Bishop, Knight, Rook, Pawn))))
  (Color (one-of (Black White))))
  (Position (row (range 1 to 8) (column (range a to h))))
  (Actions (list-of (move-procedure))))
)

```

Listing 6.3: Abstract Frame of the Entity: Chess Piece

The board configuration can be understood as a collection of relations in which pieces threaten pieces of the opposing color and defend pieces of the same color. As with entities, relations are instances of abstract concepts defined as frames. A concept schema for a binary (arity-2) chess relation would be:

```
(Relation ChessRelationA2 (ChessRelationA2-ID)
  (Name (Relation-name))
  (Kind (one-of (Offensive , Defensive)))
  (Subject (Entity-ID *))
  (Object (Entity-ID * ))
)
```

Listing 6.4: Abstract Frame of an arity-2 Relation: Chess Relation

Relations can be defensive, in which a piece protects or defends another piece of the same color or offensive in which a piece threatens a piece of opposing color. The subject and object are pointers to instances of pieces or chunks that are held as entities in the situation model. A concept schema for an arity-3 relation would be:

```
(Relation ChessRelationA3 (ChessRelationA3-ID)
  (Name (Relation-name))
  (Kind (one-of (Offensive , Defensive)))
  (Subject (Entity-ID))
  (Object_1 (Entity-ID))
  (Object_2 (Entity-ID))
)
```

Listing 6.5: Abstract Frame of an arity-3 Relation: Chess Relation

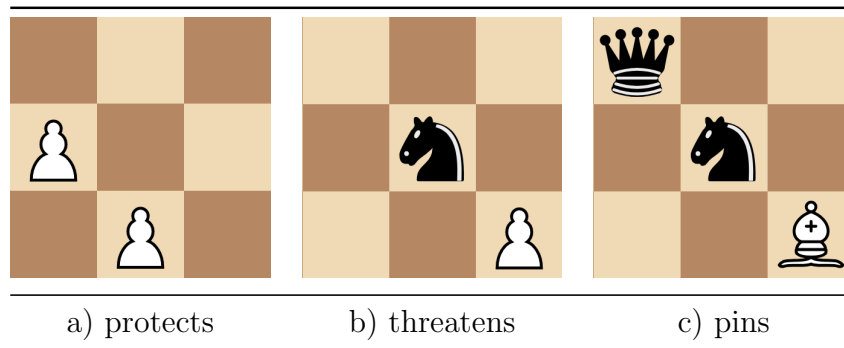


Figure 6.1: Three Examples of Relations. a) A defensive binary relation: (Pawn protects Pawn) b) An offensive binary relation: (Bishop threatens Knight) c) an offensive ternary relation: (Bishop pins Knight to Queen)

Figure 6.1 shows examples of chess relations. Figure 6.1-a is a binary defensive relation “protects”, while 6.1-b is a binary offensive relation “threatens”. Both of these can be expressed in terms of the classic binary relation schema: Subject - Relation - Object(s). Figure 6.1-c shows a ternary (arity-3) relation: “pins” with a subject and two objects.

Frames 6.6, 6.7 and 6.8 describe the relations presented in Figures 6.1-a, 6.1-b and 6.1-c respectively. Note that for every binary relation, there is an inverse relation where the Object and Subject roles are switched. For example, in Frames 6.7, the Frame *Chess_Threatens* (describing White Pawn threatening Black Knight) implies an inverse relation, called here *Chess_Is_Threatened* (describing Black Knight is threatened-by White Pawn). For arity-3 relations, there are 6 inverse relations, one each for each possible permutation of the Subject, Object_1, and Object_2 (two examples are given in Frames 6.8: *Chess_Pins* and *Chess_Is_Pinned*). Each relation constrains possible moves by the subject piece.

```
(Entity ChessPiece (WPawn.1)
  (Kind (Pawn))
  (Color (White))
  (Position ((column 2), (row 1)))
  (Actions (Move_1_Forward)))

(Entity ChessPiece (WPawn.2)
  (Kind (Pawn))
  (Color (White))
  (Position ((column 1), (row 2)))
  (Actions (Move_1_Forward)))

(Relation ChessRelationA2 (Chess_Protects)
  (Name (Protects))
  (Kind (Defensive))
  (Subject (WPawn.1 *))
  (Object (WPawn.2 *)))
```

Listing 6.6: Frames describing the binary defensive relation “Protects” (see Figure 6.1-a).

```
(Entity ChessPiece (WPawn.3)
  (Kind (Pawn))
  (Color (White))
  (Position ((column 3), (row 1)))
  (Actions (Move_1_Forward, Capture_Top_Left)))

(Entity ChessPiece (WKnight.1)
  (Kind (Knight))
  (Color (Black))
  (Position ((column 2), (row 2)))
  (Actions (Move_L_Top_Left, Move_L_Top_Right,
            Move_L_Bot_Left, Move_L_Bot_Right)))

(Relation ChessRelationA2 (Chess_Threatens)
  (Name (Threatens))
  (Kind (Offensive))
  (Subject (WPawn.3 *))
  (Object (WKnight.1 *)))

(Relation ChessRelationA2 (Chess_Is_Threatened)
  (Name (Threatened))
  (Kind (Offensive))
  (Subject (WKnight.1 *))
  (Object (WPawn.3 *)))
```

Listing 6.7: Frames describing the binary offensive relation “Threatens” (see Figure 6.1-b).

```

(Entity ChessPiece (WBishop_1)
  (Kind (Bishop))
  (Color (White))
  (Position ((column 3), (row 1))
  (Actions (Capture_Top_Left)))

(Entity ChessPiece (BKnight_2)
  (Kind (Knight))
  (Color (Black))
  (Position ((column 2), (row 2))
  (Actions (Move_L_Top_Left , Move_L_Top_Right ,
            Move_L_Bot_Left , Move_L_Bot_Right)))

(Entity ChessPiece (BQueen_1)
  (Kind (Queen))
  (Color (Black))
  (Position ((column 1), (row 3))
  (Actions (Move_Vertically , Move_Horizontally)))

(Relation ChessRelationA3 (Chess_Pins)
  (Name (Pins))
  (Kind (Offensive))
  (Subject (WBishop_1 *))
  (Object_1 (BKnight_2 *)))
  (Object_2 (BQueen_1 *)))

(Relation ChessRelationA3 (Chess_Is_Pinned)
  (Name (Is_Pinned))
  (Kind (Offensive))
  (Subject (BKnight_2 *))
  (Object_1 (WBishop_1 *)))
  (Object_2 (BQueen_1 *)))

```

Listing 6.8: Frames describing the ternary offensive relation “Pins” (see Figure 6.1-c).

6.1.3 A Situation Model for Chess

We have seen that chunks’ signatures are stored in Long-Term memory and instantiated in Working Memory to encode information from the external world or during reasoning (Miller, 1956; Minsky, 1974; Halford et al., 1998). Figure 6.2 shows the elements of our model for human cognition. With this model, Working Memory acts as a buffer for between five and nine entities. Entities represent associations of phenomena with episodic memory, procedural memory, and abstract concepts. Working memory receives information from the Attention Executive Control system that is in charge of human perception abilities (*i.e.* vision, auditory, tactile, olfactory, gustatory, etc.) and human actions that can update the environment (*i.e.* speech, manipulation, mobility, emotion display, etc.). Phenomena observed through perception abilities are stored as entities in Working Memory and will spread energy to related information in Long-Term Memory according to Hebb’s rule (Hebb, 1949) (see Section 2.3.1). Such information can be a category of chunk to encode the incoming entity, a procedural knowledge used to take action to reach a goal or an episodic memory from past experience.

Based on previous information, we can now propose a possible schema for chess situations using Frames 6.9. Entities may be instances of pieces or chunks (hierarchy of frames). The action slots provide a list of possible moves that are enabled or prevented by the situation. The emotion

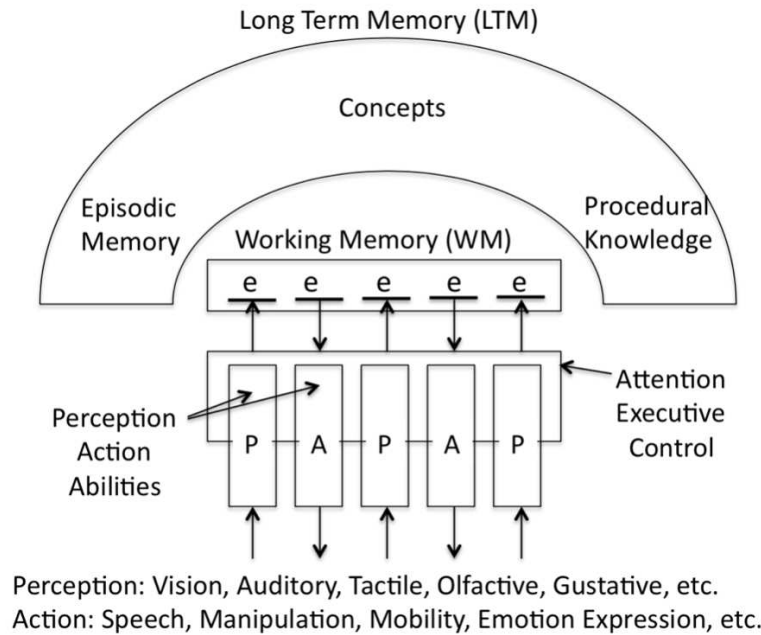


Figure 6.2: A Working Memory Model with Five Slots for Entities Guntz et al., 2018b.

slot provides values for valence, arousal and dominance acquired from experience to take into account memories encoded into memory using the special mechanism named as “Flashbulb Memory” (Brown and Kulik, 1977) (see Section 3.8 on Emotional Memory page 35). We think that emotion can not only be used for retention, but also as a guide for reasoning.

```
(ChessSituation (Situation-ID)
  (Color (one-of (Black, White)))
  (Relations ([Relation-ID, ...]))
  (Entities ([Any-of (Piece-ID, Chunk-ID), ...]))
  (Moves ([Move, ...]))
  (Emotion (Valence, Arousal, Dominance))
)
```

Listing 6.9: Chess Situation described with Frame.

The syntax $[Expr, \dots]$ indicates a facultative element that can be a list of one or more elements. The number of entities and relations within a situation is limited by the number of elements in working memory. This number should not exceed Miller’s limit (Miller, 1956, see Section 2.3.2) and is likely to be much smaller as part of working memory will be consumed by other information about the game context such as imminent threats or time constraints. For example, when considering the consequences of a move, the player must retain the current situation, the candidate move, and the resulting situation in working memory, as well as other information about the game context. This would seem to limit the game situation model to 3 or at most 4 entities. For example, a typical situation model may be composed of one defending chunk, one attacking chunk, and a single active piece.

6.2 Evidence for Awareness from Observation

6.2.1 Computation of the Evidence from Observations

Fixations and emotions provide evidence for awareness of chess chunks. In this section, we propose a mathematical model for collecting such evidence. As described in chapter 4, the retina of the human eye is composed of a dense central region (a fovea) surrounded by a larger peripheral region (the parafovea). We have designed our chess display such that each chess square corresponds roughly to the size of the region of space that projects onto the fovea. Thus we can represent our chessboard as an 8 x 8 grid of discrete fixation positions.

Chess chunks have a spatial extent that can be described in terms of chess squares. Each individual piece occupies a square, threatens a set of squares, and can potentially move to a set of squares. This set of squares is the “*Spatial Extent*” of the piece. Fixating any of these squares is evidence for awareness of the piece.

In addition, pairs of pieces form compound chunks that cover the individual squares of the two pieces, plus the squares between them. The resulting chunk has a spatial extent that is the union of the spatial extent of its components. Chunks can be composed hierarchically, with each chunk having a spatial extent composed of the union of the spatial extent of its components. For each chunk, we can construct a list of squares in its spatial extent. We can use the spatial extent of the chunks to build a form of index for each square. For each square, we construct a list of chunks that include that square in its spatial extent. Thus we have a 8 x 8 grid of squares with a list of chunks for each square. When the subject fixates a square, this can be counted as evidence for any of these chunks.

In the absence of knowledge of the likelihood of awareness of chunks, we can assume that all chunks are equally likely. Thus if a square points to N chunks, we can assume a probability of awareness of $1/N$ for each chunk. We ignore squares that do not point to any chunk.

Assume a set of K chunks $\{C_k\}$ and J fixations $\{F_j\}$. In the absence of information, the probability of awareness of any chunk is :

$$P(C_k) = \frac{1}{K} \quad (6.1)$$

A fixation F_j gives evidence for awareness of a chunk C_k if the position of F_j is on a square, S , which is included in the Spatial Extent SE_k of C_k . In that case, we write: $F_j \in S$ and $S \subset SE_k$. For better readability, we can omit the information of the square and directly write the following: $F_j \subset SE_k$.

Let N_s be the number of chunks that include a square S in their Spatial Extent SE . The probability of awareness of a chunk C_k given a fixation is F_j at square S is:

$$P(C_k|F_j \subset SE_k) = \frac{1}{N_{s \subset SE_k}} \quad (6.2)$$

If, on the other hand, a fixation F_j is not included in the Spatial Extent of C_k gives, then no information of the awareness of such chunk is given:

$$P(C_k|F_j \not\subset SE_k) = 0 \quad (6.3)$$

After J fixations, the probability of awareness of each chunk C_k is obtained using the law of total

probability:

$$P(C_k) = \sum_{j=1}^J P(C_k|F_j) \cdot P(F_j) \quad (6.4)$$

The terms of the sum in equation 6.4 are null if the fixation $F_j \not\subseteq SE_k$. Thus, we can be simplified using only the terms that are not null. Let consider the set $\{F_i\}$ a subset of $\{F_j\}$ that verifies for every fixation: $F_i \subseteq SE_k$, then the previous equation can be written again as:

$$P(C_k) = \sum_{i=1}^I P(C_k|F_i \subseteq SE_k) \cdot P(F_i \subseteq SE_k) \quad (6.5)$$

The second term of the sum computes the probability that a fixation is made on a square that is included in the Spatial Extent of chunk k . The probability of a fixation being on a square depends on many variables: its relative distance to the center if a piece is on the square, its surroundings, etc. The player expertise also impacts this probability: as we have seen in Section 4.1.4, experts tend to fixate between pieces, using their parafovea vision to gather information, while novices do fixations more on pieces. For the early development of this model, we use a naïve approach that considers that all square have the same probability of being fixated: $P(F_i \in S) = 1/64$. Finally, this probability is proportional to the number of squares contained in SE_k (denoted here $\#SE_k$).

$$P(F_i \subseteq SE_k) = \frac{\#SE_k}{64} \quad (6.6)$$

Equation 6.5 can be simplified as:

$$P(C_k) = \frac{\#SE_k}{64} \cdot \sum_{i=1}^I P(C_k|F_i \subseteq SE_k) \quad (6.7)$$

6.2.2 Evidence for Chunks from Fixations

Evidence can be deduced from players' visual attention and more precisely from fixations (see Section 4.1.3). For that, we set a counter of activation for every chess chunks that get activated when a fixation is detected inside his visual span. As a piece or a square can belong to different chunks, a single fixation may activate more than one chunk. A fixation does not imply that the activated chunks are actually present in the player's WM, however, frequently fixated chunks are considered to be more likely to be present in WM. Once all fixations have been considered, chunks' evidence can be computed by summing all activations and using definitions from section 6.2.1.

Let's consider the following chunks, displayed in Figure 6.3, as examples. On this image, two different chunks are overlapping (**i.e.** they share common pieces and squares):

- Chunk 1: White Bishop protecting White Pawn, let's name it: CH_PROT;
- Chunk 2: White Bishop threatening Black Queen, let's name it: CH_THREAT;

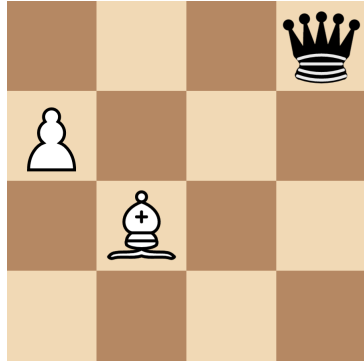


Figure 6.3: Toy Example with 3 Pieces.

```

(Piece (WBishop)
  (Kind (Bishop))
  (Position (b2))
  (Actions (...)))

(Piece (BQueen)
  (Kind (Queen))
  (Position (d4))
  (Actions (...)))

(Piece (WPawn)
  (Kind (Pawn))
  (Position (a3))
  (Actions (...)))

(Relation (R1)
  (Name (Protects))
  (Kind (Defensive))
  (Subject (WBishop))
  (Object (WPawn)))

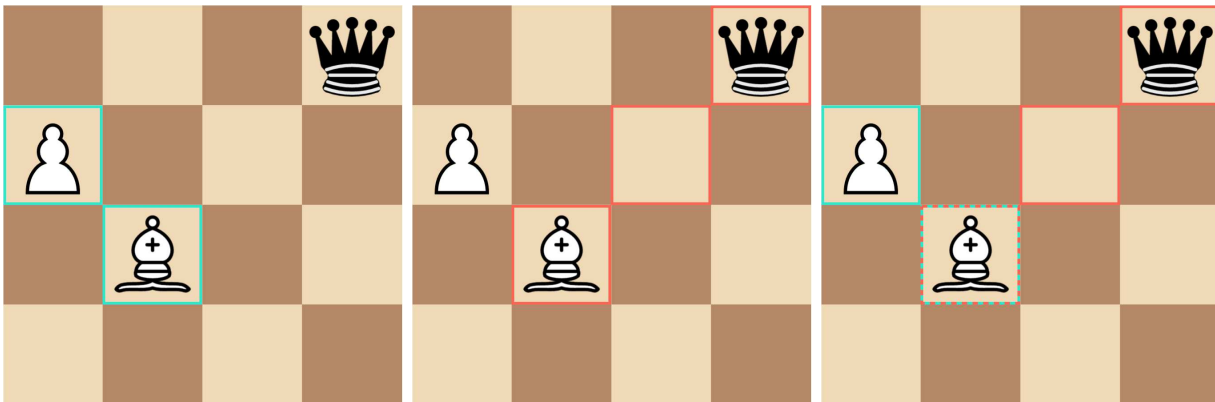
(Relation (R2)
  (Name (Threatens))
  (Kind (Offensive))
  (Subject (WBishop))
  (Object (BQueen)))

```

Listing 6.10: Pieces and Relations presented in Frames.

The visual spans for each chunk are displayed 6.4a and 6.4b. CH_PROT covers two squares; each centered on both the subject (Bishop) and the object (Pawn) of the chunk. About the second chunk, CH_THREAT, three squares are involved. These squares include the diagonal between the White Bishop and the Black Queen, as well as the Bishop and Queen.

When a fixation is recorded, the activation for all of the chunks whose spans overlap the fixation are increased. For example, take a scan-path with three fixations F1, F2 and F3 as examples, displayed in Figure 6.5. Fixations from this example cover three cases:



(a) Visual Span of Chunk 1: CH_PROT (b) Visual Span of Chunk 2: CH_THREAT (c) Visual Span of both Chunks with a common piece: Bishop

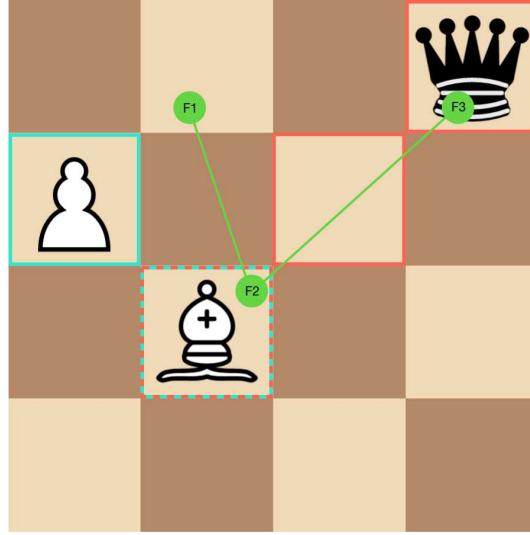


Figure 6.5: Scan-Gaze Example with Three Fixations: F1, F2 and F3.

- F1 is not included in any visual span, we consider that no chunk is activated,
- F2 is in both visual span: in CH_PROT and CH_THREAT, we say that both chunks are activated and considered by the player,
- Finally, F3 is only involved in one chunk: CH_THREAT, we imply that only this chunk is activated.

Once the full eye-scan is analyzed, evidence counters for each chunk can be computed. Let's consider the set $\{C_k\}$ composed of two chunks $\{C_1, C_2\}$. C_1 being CH_PROT and C_2 CH_THREAT. With no information, we assume that each chunk has the same probability of being observed:

$$P(C_k) = \frac{1}{K} = \frac{1}{2} \quad (6.8)$$

Three fixations F1, F2, and F3 are recorded. If we consider each fixation now. F1 is not included in any Spatial Extent of chunks. Thus this fixation provides no evidence:

$$\begin{aligned} P(C_1 | F_1 \notin SE_1) &= 0 \\ P(C_2 | F_1 \notin SE_2) &= 0 \end{aligned}$$

The second fixation F2 is recorded on the square of the Bishop. This square is involved in the Spatial Extent of both chunks: C1 and C2, thus $N_s = 2$ and:

$$\begin{aligned} P(F_2 \in SE_1) &= \frac{\#SE_1}{64} = \frac{1}{32} \\ P(F_2 \in SE_2) &= \frac{\#SE_2}{64} = \frac{3}{64} \\ P(C_1 | F_2 \in SE_1) &= \frac{1}{N_s} = \frac{1}{2} \\ P(C_2 | F_2 \in SE_2) &= \frac{1}{N_s} = \frac{1}{2} \end{aligned}$$

Finally F3 is focused on the Black Queen inside the Spatial Extent of C2:

$$\begin{aligned}
 P(C_1 | F_3 \notin SE_1) &= 0 \\
 P(F_3 \subset SE_2) &= \frac{\#SE_2}{64} = \frac{3}{64} \\
 P(C_2 | F_3 \subset SE_2) &= 1
 \end{aligned}$$

Applying now the total law of probability defined for the evidence in equation 6.7, we can write the probability of awareness of C1:

$$\begin{aligned}
 P(C_1) &= \frac{\#SE_1}{64} \cdot \sum_{i=\{2\}} P(C_1 | F_i \subset SE_1) \\
 P(C_1) &= \frac{1}{32} \cdot P(C_1 | F_2 \subset SE_1) \\
 P(C_1) &= \frac{1}{32} \cdot \frac{1}{2} = \frac{1}{64}
 \end{aligned}$$

And the same can be applied for C2:

$$\begin{aligned}
 P(C_2) &= \frac{\#SE_2}{64} \cdot \sum_{i=\{2,3\}} P(C_2 | F_i \subset SE_2) \\
 P(C_2) &= \frac{3}{64} \cdot [P(C_2 | F_2 \subset SE_2) + P(C_2 | F_3 \subset SE_2)] \\
 P(C_2) &= \frac{3}{64} \cdot \left[\frac{1}{2} + 1\right] = \frac{9}{128}
 \end{aligned}$$

To conclude this toy example, we can say that the probability of awareness of chunk C2 is higher than C1 and the individual who performed the fixations have higher chance to hold C2 in his Working Memory than C1.

6.2.3 Evidence for Experience from Emotions

Our hypothesis is that while engaged in problem-solving, chess players generate and explore a tree that combines possible interpretations of the situation and possible reactions. This exploration is influenced by past experiences with similar situations encoded as affect (Valence, Arousal, and Dominance) in LTM. This affect is displayed as facial expressions as the player explores the tree of possible interpretations. From what we have learned with our first experiment (See Chapter 5), these signals are stable most of the time, except for a few peaks. We suggest that these peaks are unconscious somatic reactions correlated to their exploration of possibilities. Analyzing player's valence, encoding either pleasure or displeasure, and arousal for threats and opportunities could give us valuable information on their current evaluation of the game.

Let us consider an example with two players of different levels, a novice, and an expert, analyzing a small situation displayed in Figure 6.6. This toy chess situation illustrates a White Bishop protecting a White Pawn while at the same time threatening the opponent's Black Queen. In an effort to focus solely on emotions, we consider in this example that both players realized the same displayed fixations on the board, with the same timing. That being said, let us imagine how the novice would mentally react if he were playing Black (Note: in chess, the Queen is the second most important piece, the first being the King). Figure 6.7 is the tree of possible moves

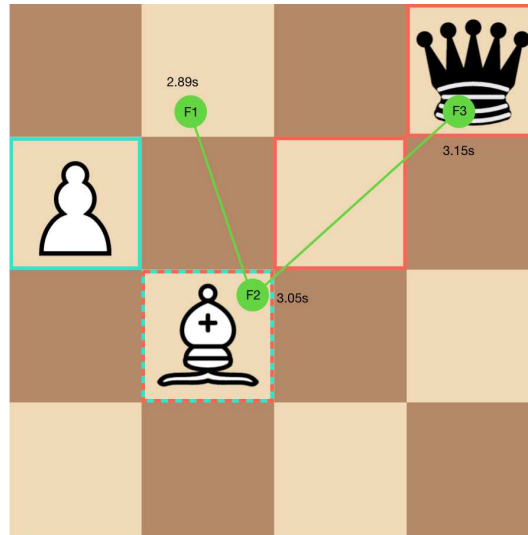


Figure 6.6: Fixations Example with Time of Apparition in Seconds.

considered by the novice player (fictional scenario for illustration purposes). On top of the tree is the situation S_0 displayed on the board: the Queen is threatened. From S_0 , two situations are accessible. The first one is $S_{1.A}$, reached by taking Move A, where the Queen is taken, and Black find themselves in a material disadvantage. The second situation is $S_{1.B}$ accessible from Move B; the Queen has been secured. According to our hypothesis, each mentally evaluated moves are also linked to previous and similar encountered situations. For each situation, Valence and Arousal are plotted in blue and red, respectively. The situation S_0 as it stands displays an incoming threat for the Queen, and it should be apparent on the arousal plot. Small displeasure should also be visible. We suggest that players would consider first moves that maximize Valence (*i.e.* pleasure) and Arousal (*i.e.* opportunity). Then players would first evaluate moves where the Queen is safe as it removes threat and displeasure. Novice players should consider move B.

Nonetheless, losing the Queen does not have to mean defeat. Some advanced strategies require a sacrifice of a strong piece as bait in order to take a significant advantage. Players need experience to dare to play these moves. Let us consider an expert playing Black, its tree of possibilities is showed in Figure 6.8 (fictional scenario for illustration purposes). We notice first that he evaluates a move further (2 moves from S_0) than the novice. This does not imply that experts evaluate deeper tree of possibilities than novices; it has been proven several times through researches (De Groot, 1978; Gobet and Simon, 1996; Simon and Chase, 1988). However, this does mean that the affect retrieved from Long-Term memory of the expert about the situation where he loses his Queen is not always negative. During hours of practice, advanced players have occasionally won games using a Queen sacrifice, retaining positive affect from such situations. This is visible on the situation $S_{2.B}$ where, indeed, an advantage is taken, maybe by threatening the opponent's King. It is also possible that sacrificing the Queen represents a material disadvantage with no positive outcome. In this case, both the expert and the novice react the same way, as shown in $S_{2.A}$.

Finally, we may also hypothesize which chunks are being considered by players by combining the emotional response, eye-scan, and temporal information. If we consider the scan-path displayed in Figure 6.6, we can correlate the peak of arousal with the fixation and deduce that the chunk related may have triggered the emotional response recorded.

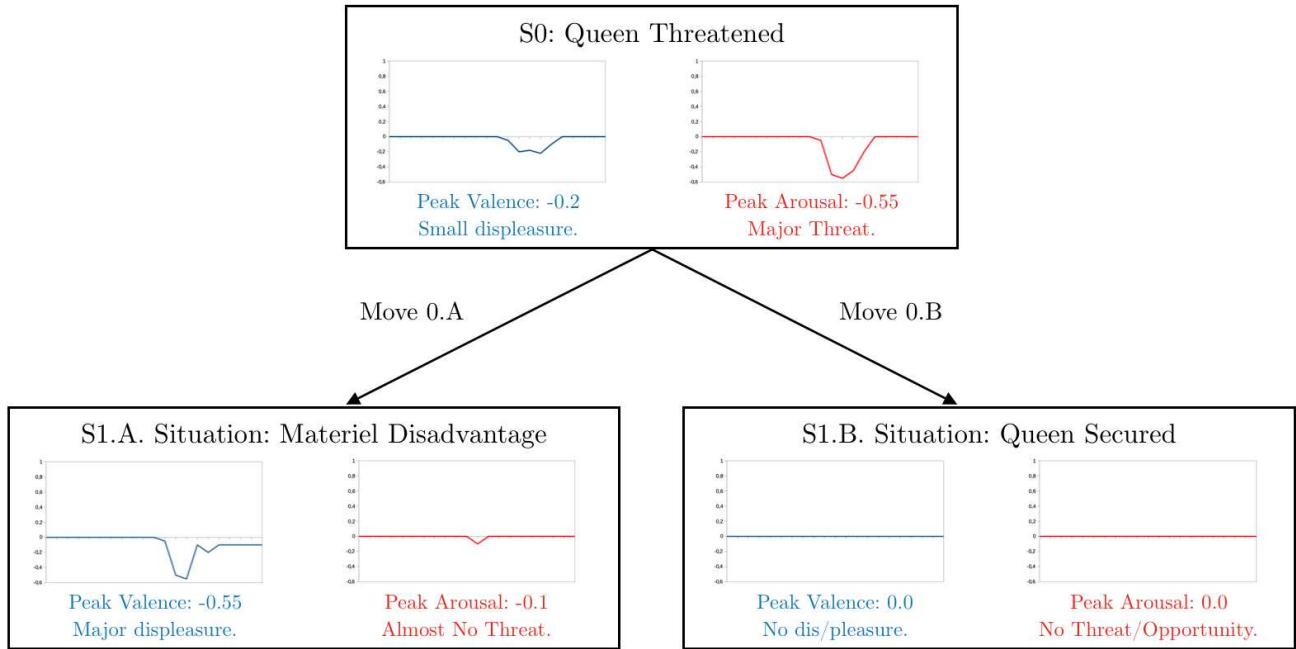


Figure 6.7: Tree of Possible Moves with Recalled Affect from Long-Term Memory. Novice’s Point of View. (Fictional Scenario for Illustration Purposes)

6.3 Conclusion

In this chapter, we propose a model for representing awareness from observation of eye-gaze and emotions. We employ frames to represent concepts and chunks manipulated by chess players. We then illustrated the use of Frames by representing several chess chunks, simple as complex, proving the ease of use of this framework.

Based on the literature (see Section 2.2), we also propose a model to describe the content of Working Memory and its interactions with other cognitive entities (Long-Term Memory, human perception abilities, etc.). The elements manipulated in Working Memory and Long-Term Memory are represented by Frames.

Evidence for awareness of chunks is provided by recordings of eye-gaze, emotion, body movements. While we can not conclude with certainty which chunks are attended, we have shown a model that makes it possible to accumulate evidence for awareness.

In order to develop our model, we have conducted a second experiment in which we players were asked to explain their reasoning.

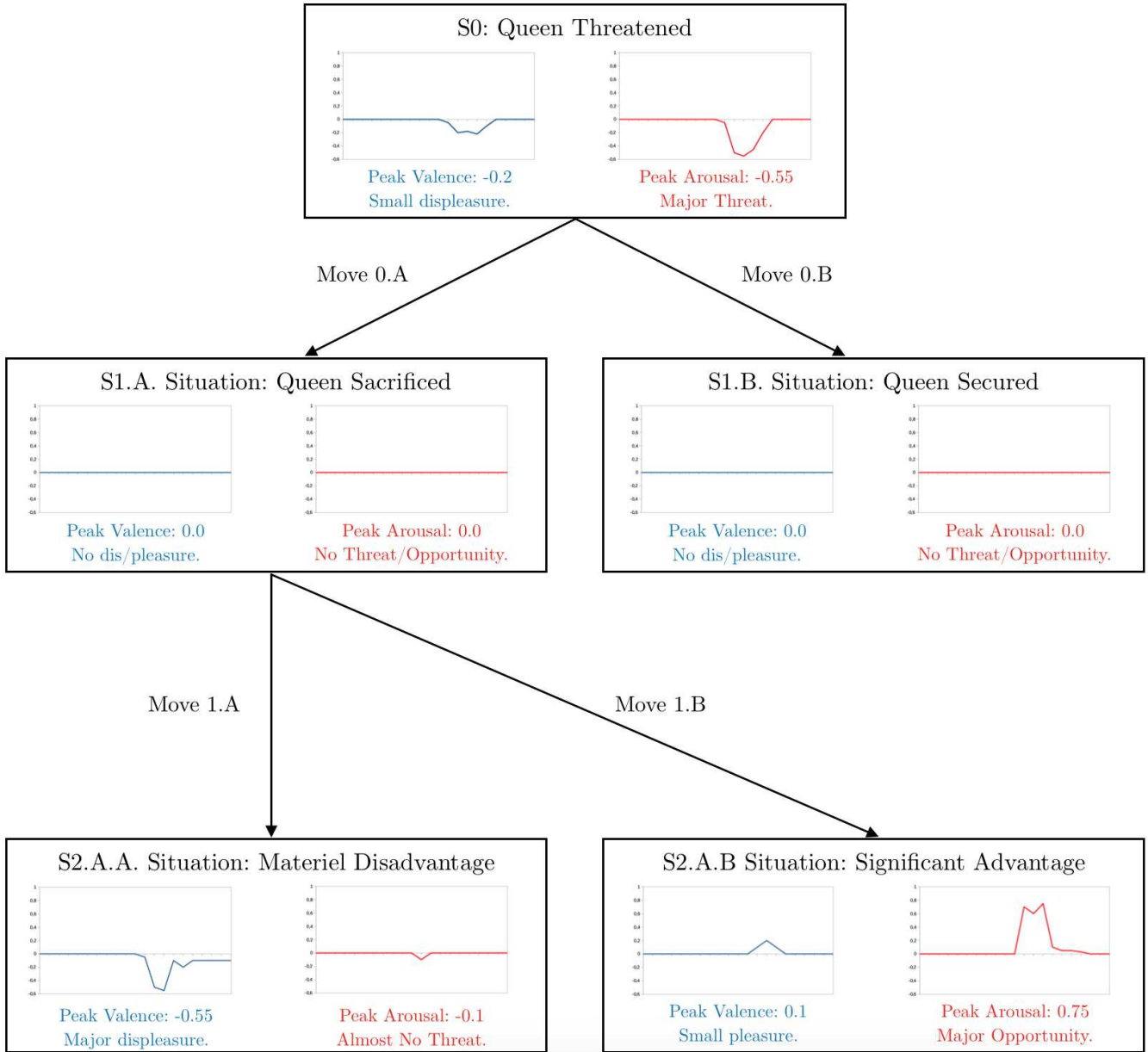


Figure 6.8: Tree of Possible Moves with Recalled Affect from Long-Term Memory. Expert's Point of View. (Fictional Scenario for Illustration Purposes)

Chapter 7

An Experiment to Observe the Role of Emotion in Reasoning

To verify the model described in Chapter 6, we devised a new experimental protocol, in which players are asked to describe their reasoning after attempting to solve each problem. Indeed, to quote again Simon and Barenfeld: “*Records of eye movements can only show the succession of fixations; they cannot show precisely what information is being processed at each moment*” (Simon and Barenfeld, 1969). By adding a verbal protocol, the principal motivation is to identify the chess chunks with which the players reason and to correlate these to chunks that are indicated by fixation. However, a player may display over 100 fixations while seeking a solution to a problem. To focus on the most important fixations, we refine our search for information based on a lesson learned from our first experiment: emotions are correlated with the situation understanding. We thus anchor our analysis to a few seconds before and after strong peaks in emotion.

This section begins with a review of the literature Verbal Analysis (Section 7.1) followed by a description of a pilot experiment (Section 7.2).

7.1 Study on Self-Reporting

Introspection or the “examination” of an individual’s conscious thoughts and mental images has long been used in psychology as a tool to uncover information about awareness and reasoning (Ericsson, 2006). Verbal Analysis is widely used in studies about Learning. Verbal Analysis seeks to describe how individuals encode and verbalize information (Balzarini, 2013). This is performed by encoding a verbalized idea with a unique element called a “*verbatim*”. Verbatims from several individuals can be correlated, with statistics over several individuals used to describe the effectiveness of a presentation for learning, or for differentiating knowledge of experts from novices.

One popular method of Verbal Analysis, attributed to Newell, Simon, et al. (1972), is called “*Protocol Analysis*” which consists of asking individuals to “think-aloud” every piece of information that comes to their mind while engaged in problem-solving. This approach is systematic and does not let time for the individual to explain or justify its verbalized thoughts. This method aims to identify which solving procedures (**i.e.** series of tasks and operators) used by the subject. Gathered information provides insights on the elaborated strategy in terms of a sequence of operations. Uncovering strategies is valuable information as studies have shown that the definition of the problem guides experts’ strategy, whereas novices’ focus is goal-oriented (Simon and Simon, 1978). Both views are composed of a different sequence of operations.

Another approach named “*Verbal Data Analysis*” requires the subject to rationalize and explain actions as they are solving the task, raising the risk of interfering with performance (Chi, 1997). In contrast with *Protocol Analysis*, the *Verbal Data Analysis* retraces the subject’s resolution path without knowing in advance the possible relevant concepts or operators. No information on the chosen strategy is given, but this method identifies the set of knowledge involved by a subject in solving a problem.

7.1.1 Verbal Analysis in Chess

In his extensive work, De Groot gathered verbal analysis of both experts and novices players (De Groot, 1978). He presented to players representative games that required immediate action (*i.e.* play the next move) while instructed them to verbally report their thoughts (think-aloud protocol, Ericsson, 2006). From these verbal protocols, De Groot structured the selection of the next move in different phases (the Orientation Phase, the Exploration Phase, the Investigation Phase and the Proof Phase). He showed that both experts and novices first familiarized themselves with the presented position and, in a second time, explored salient aspects of the position such as potential attacks or defenses. It is only afterward that players analyzed the validity of the possible moves by exploring consequences and evaluating resulting positions. De Groot stated that the lack of verbalization of the best movement during the exploration phase by novices is a consequence of an inferior position representation created during the first phase. In contrast, experts were inclined to verbalize strong moves even during the initial phase. To explain this discrepancy in representation, De Groot suggested two mechanisms developed by experts to help them find and select moves:

1. Rapid perception and encoding relevant structures in the chess position;
2. Superior performance in planning help to generate potential new and better moves.

An example of verbal protocol from a good and expert chess players is given Figure 7.1 (Ericsson and Charness, 1994). These verbal analysis of chess players suggest that the extensive knowledge of experts are not the only responsible of their performances. The ability to encode and manipulate internal representations of chess positions is essential. This can be observed with chess masters that can play blindfold chess (without board and pieces), an activity that is impossible for less-skilled players. Internal manipulation of chess representation allows experts to anticipate the consequences of possible next moves and to consider future threats (Ericsson and Charness, 1994).

7.1.2 The Concurrent and Retrospective Approaches

Verbalization of thoughts can be done following two approaches: during decision making (*i.e. concurrent data*) or after task completion (*i.e. retrospective data*). Although both methods are used to reveal decision-making processes, significant differences exist between the two methods, and the choice of one or the other should be made accordingly to the objectives to be achieved and the environmental experiment (Kuusela and Paul, 2000).

Concurrent data are collected during task completion. “*Protocol Analysis*” (Newell et al., 1972), or think-aloud protocol, are considered as concurrent verbal methods as they instructed subject to verbalize their thoughts during problem-solving. These protocols report information from ongoing cognitive activities and the way they are processed as they go along. A characteristic of concurrent analysis is the rich collection of collected data. This method is an appropriate and recommended method to reveal subject’s thought processes about features and procedures used

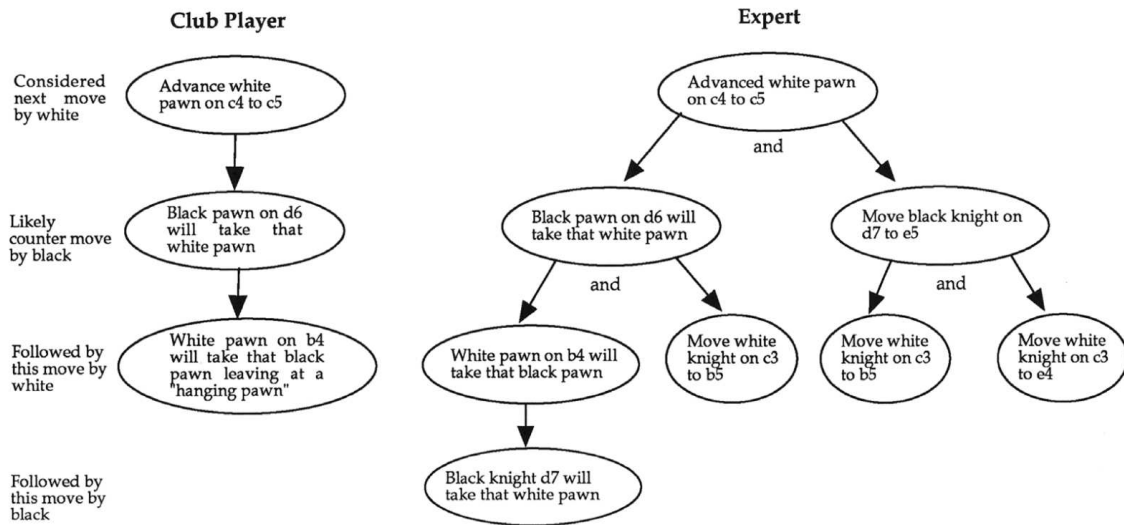
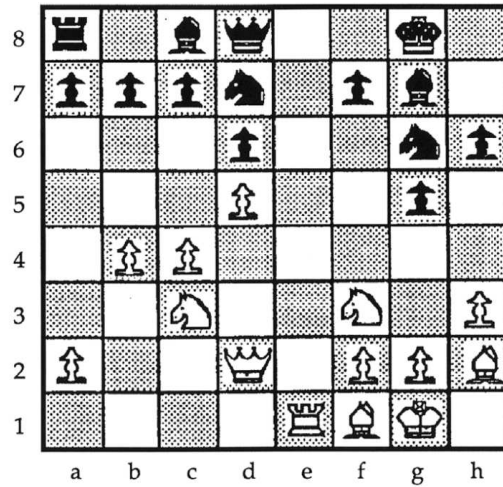


Figure 7.1: Example of Presented Chess Position to a Good Club Player (ELO = 1657) and to an Expert (ELO = 2004). Instructions are to select the best move for White. The respective think-aloud protocol is shown to illustrate differences in both evaluating and planning a specific move: *White Pawn is moved from C4 to C5* (which is the best move for this position). (Source: Ericsson and Charness, 1994)

during decision making. However, a disadvantage arising from this phenomenon is the time-consuming task to encode and process all generated data. Finally, task performance may also be impacted by verbalizing thought processes. If instructed to explain their thoughts while performing the task, subject's ability for learning tend to improve (Chi, 1997), where no effect has been shown for a simple Think-aloud verbalization (no explanation).

Retrospective data are gathered after task completion. Ideally, these are collected immediately after the task is completed, while information is still in an accessible state in subject's short-term memory. As one may expect, the content of subject's short-term memory after task completion is concepts related to the final task decision. However, during problem-solving, participants typically explore many strategies before considering their final decision. During retrospective verbalization, participant tends to focus on their final choice and will experience great difficulty in retracing intermediates steps (Ericsson and Charness, 1994). Moreover, verbatim encoding encountered difficulties in solving the task are fewer in retrospective condition. A possible explanation is that participants tend to mention their problems while facing them (Kuusela and Paul, 2000). The

main advantage of using retrospective conditions is not to alter subject’s performance. In contrast with concurrent analysis, no cognitive resource is consumed to verbalize thoughts. Thus, if the study focuses on the final decision without altering participants’ natural behavior, retrospective protocols should be considered over concurrent ones.

7.1.3 Summary

Table 7.1 summarizes the reviewed literature from previous sections about Verbal Analysis procedures (**i.e.** Think-Aloud and Explanations) and specific conditions (**i.e.** Retrospective and Concurrent) along with their advantages and disadvantages (Simon and Simon, 1978; Chi, 1997; Kuusela and Paul, 2000; Ericsson, 2006; Balzarini, 2013).

	Retrospective Condition (i.e. After task completion)	Concurrent Condition (i.e. During task)	Advantages (Analysis)	Disadvantages (Analysis)
Protocol Analysis (Think-Aloud)	Think-aloud after task completion	Think-aloud while solving the task	Generate many verbatim	Data processing is time consuming
Verbal Data Analysis (Explanations)	Explanations after task completion (RTE)	Explanations of considering strategies (incoming thoughts) while solving the task (CVA)	Identify relevant knowledge used to take decision	Require cognitive resources
Advantages (Condition)	Keep an ecological test environment	May improve performance (learning) Intermediates steps are vocalized Facing problems are vocalized		
Disadvantages (Condition)	Focus is done on final decision Intermediates steps are barely verbalized Faced problems are barely verbalized	Alter Reaction Time Part of Cognitive resources are allocated to vocalize (and not for problem-solving)		

Table 7.1: Summary of Verbal Protocols and Conditions with their Pros and Cons. The two bolded protocols are the ones tested in our pilot experiment.

In view of understanding which concepts are used by chess players to solve the task and, most of all, which considered concepts are arousing affect, a pilot experiment has been conducted involving a subset of methods (RTE and CVA) to observe to find out which one best meets our expectations.

7.2 A Pilot Experiment

In March 2018, we conducted a pilot experiment to gather information on two Verbal Analysis protocols:

1. **Concurrent Verbal Analysis (CVA)**: players are instructed to verbalize and to **explain** their strategies and thoughts **while** completing the task.
2. **Retrospective Task Explanation (RTE)**: players are instructed to **explain, after** task completion, strategies they consider to take their final decision.

We emphasized that the two proposed Verbal protocols are not considering as Think-Aloud protocols, where participants would vocalize their incoming thoughts without explaining them. Explanations are the most relevant method to gather concepts considered in subjects’ strategies (Chi, 1997; Balzarini, 2013).

7.2.1 Procedure

This pilot experiment is composed of eight N-Check-Tasks of increasing difficulty: 4 tasks are taken from the first experiment (see Section 5.3.3), and the 4 remaining are new ones specially designed. The initial modalities are still recorded (Body, Gaze, Facial). Out of those eight chess puzzles, some of them should be explained by the subjects, such as:

1. The 2 first tasks are not explained;
2. The 3 following tasks are explained using one method (RTE or CVE);
3. The 3 remaining tasks are explained using the other method (not chosen in 2).

The initial choice of the RTE or CVA method for step 2 was made randomly among the participants to control the impact of the first applied method on the second. Overall, the selected tasks are easier than the ones from experiment 1. That way, players of all levels could engage themselves in problem-solving without being discouraged in advance. The goal of this pilot experiment is to collect verbal data (verbatim); discouraging participants would thwart our efforts.

Colleague chess players from our laboratory volunteered to participate in this pilot experiment. Upon arrival, participants are first being instructed that this experiment is performed to analyze the reasoning of chess players, based on external observation. In order not to bias the behavior of the participant during the experiment, little information about the interest of the experiment is given. The participant has been informed, at his or her discretion, of the project's stakes and research hypotheses once the experiment has been completed. Secondly, information on the nature of the tasks to be solved is given: chess problems, similar to puzzles found in magazines or on the web, are presented one after the other. For each problem, the examiner announces the played color, however, in contrast with the first experiment, no further information is delivered on the goal to achieve with the chess task. Participants choose to play either offensively or defensively. Then, the participant is invited to take as much time as needed to analyze the problem before playing the move(s) that seem(s) most appropriate to the situation.

For some tasks, participants were asked to explain the steps he or she goes through before making the decision to make a move, consistent with “**Concurrent Verbal Analysis**” (CVA). To do so, he must rationalize the characteristic elements of the problem (pieces, positions, configurations, etc.) that he may or may not consider as relevant during his analysis and decision making. For tasks requiring a “**Retrospective Task Explanation**” (RTE), identical instructions as CVA procedure were given, however, participants were asked to verbalize their explanation once the task had been completed. Instructions about eye tracking calibration and the posture to adopt (e.g. avoid large movements) are finally given.

7.2.2 Results and Discussions

In total, 4 of our colleagues took part in the pilot experiment, two of them were casual players (novices with no ELO ranking) while the other two had intermediate levels (ELO ranking around 1400). None of them were playing in chess clubs at the time of the experiment. Also, we asked two participants to provide their RTE right after the task completion (i.e. before moving on the next task), whereas the two other participants were asked to complete all tasks before coming back on them to provide their RTE.

Verbal protocols were recorded using the Kinect microphone. These protocols were then played back and transcribed by hand. This experiment provided valuable information on the potential adequacy of these procedures for our experiments.

Observations on CVA

Regarding the Concurrent Verbal Analysis, we identified several advantages: many concepts are verbalized by the participants providing a rich amount of data. Among these verbatims, most (if not all) of the key concepts involved in the final decision of the participant are mentioned. As

participants explain their thoughts while considering different strategies, intermediate steps are present in their verbal reports. Other concepts, less relevant for completing the task, are also verbalized, but quickly discarded. These advantages are in line with the literature; however, some drawbacks raised some concerns: the participants' speeches were very close to a Protocol Analysis (Think-Aloud), and some verbalized thoughts were not explained. Moreover, participants do not behave naturally when giving explanations. Their emotional reactions and eye trajectories are not those we would observe in an ecological context.

Observations on RTE

About the Retrospective Task Explanation, we noticed some advantages: some key concepts used in the participant's decision-making are verbalized. The process keeps the test environment ecological and it is not impacted by the verbalization; thus we do not notice any impact on performance. Various inconveniences have been observed based on whether RTE is performed right after task completion or only when all tasks have been completed: The first case is when the RTE is performed after all tasks have been completed, participants expressed great difficulties in remembering both the task and their choices. As mentioned earlier, this is mostly due to the content of STM being replaced over time with fresher information (Kuusela and Paul, 2000). About the second case, when the RTE is performed right after the task, most participants focus the speech on their final decision, explaining concepts involved in this decision. Thus, we notice that intermediate steps and the strategies that were studied before being rejected are barely mentioned. Last but not least, the focus of verbalized concepts is highly influenced by the opponent's moves (played automatically, by a Stockfish algorithm, once a move is made by the participant). This is especially true when the opponent's responses are not expected by the participant or when he realized that he played a bad move. In this situation, some participants were reluctant to explain their final decision, as they realized it was not a good solution.

Discussion

This pilot experiment brought valuable information on both verbal protocols. While CVA provides a rich amount of data about intermediate steps and the final decision, it compromises the natural behavior of chess players. Also, some cognitive resources are allocated to verbalize thoughts and are thus not used in problem-solving. RTE presents significant problems with the report content. Concepts that are not directly involved in the final decision are barely mentioned, and the speech can be highly influenced by the opponent's counterattacks. Nevertheless, RTE assures that chess players' behavior is natural. This would offer us the opportunity to interpret their eye trajectories and physiological responses synchronously with their verbal reports. The timing plays an important role. Indeed STM decays over time and gets refreshed with new incoming information. RTE should be realized as soon as the task is completed to maximize valuable information from the verbal analysis.

Taking all these observations into account, we decided to rely on the Retrospective Task Explanation (RTE) protocol to report verbal analysis of chess players right after task completion. We conclude this review on self-report by mentioning two essential points for examiner that would interview participants: Michelene Chi, in (Chi, 1997), emphasized "*How the experimenter should be as unintrusive or as uniformly intrusive as possible*" to avoid any influence and ensure consistency between participants; she also raised the following problem: "*How to control for the fact that some people are more verbose than others?*". This concern must be carefully thought out

in advance. Chi stated that if a subject takes the time to explain an idea, so it is clear that this idea is of significant value to him and should be considered as such.

7.3 A Second Experiment

7.3.1 Objectives and Research Questions

The objective of this experiment is to validate if evidence of awareness can be observed using our cognitive model described in the previous Chapter. From these pieces of evidence, several hypotheses were tested:

- Could evidence of eye-gaze, valence, arousal, and frustration be correlated with the four phases of reasoning proposed by De Groot (1978)?
- Could the content of the Working Memory of chess player be inferred using their Verbal Protocol?
- Could emotional evidence be better interpreted using the content of Working Memory?

7.3.2 Participants

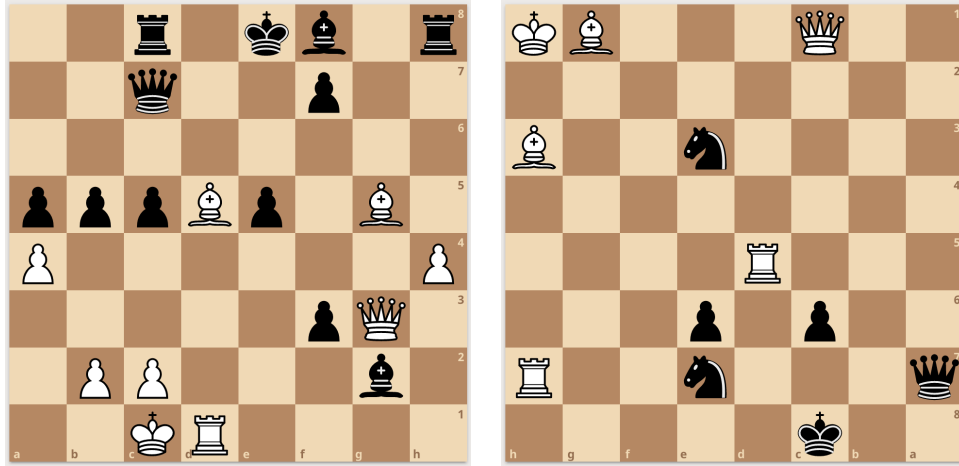
For our second experiment, the same announcement as the past year has been passed along chess clubs, local universities, research laboratories, and within the metropolitan area. The experiment has been held in the IMAG building situated on the campus of Saint-Martin-d'Hères. We recruited 23 subjects (12-55 years, 4 females, age: $M = 29.17$; $SD = 10.56$), consisting of 2 experts with ELO rating 1930 and 2000, 19 intermediates rated between 1197 and 1700 ($M = 1417.05$; $SD = 102.65$) and 2 unranked beginners. Twelve of the intermediates were casual players who were not currently playing in clubs.

7.3.3 Designing Chess Tasks

This time again, we rely on the work done in collaboration with the CITEC team of Bielefeld to design the chess tasks. For the first experiment, two kinds of tasks were designed: Openings and N-Check-Mate. Unfortunately, we noticed that openings situations were not suitable for our study, due to players being too over-familiar with this particular stage of the game and they played automatically without initiating any cognitive process. As a replacement, we designed a particular type of task that aims to elicit emotion: *Survival tasks*. In addition, new N-Check-Mate tasks have been designed, and some tasks used in the first experiment have been reused. Figure 7.2 shows example of designed tasks.

Survival Tasks. Three tasks have been elaborated that place participants in hopeless situations. These situations are meant to elicit strong affect as the more the participant assesses the situation, the more the hopelessness of the situation becomes apparent. When presented to experts to validate the tasks, most told us they would rather give up because there was no point in trying.

N-Check-Mate. These tasks are the same as designed in Experiment 1 (see Section 5.3). They are defined, such as one easy task (mate in 1), two tasks with medium difficulty (mate in 3), and one hard task (mate in 5). Out of these four tasks, two of them present unbalanced situations in favor of the opponent, forcing conservative players to focus on defensive moves. The most attentive and experienced players will, however, quickly find the solution to solve the game.



(a) N-Check-Mate Task with $N=5$. White to Play. (b) Hopeless Situation. Black to play (i.e. Black to survive).

Figure 7.2: Example of Designed N-Check-Mate (left) and Survival Task (right)

7.3.4 Ontology of Chess Concepts

For each elaborated task, we extracted all relevant concepts that may be considered by chess players. Every piece, relevant relations (n-arity), and configurations of pieces are listed. Relevant concepts have been listed in collaboration with the CITEC team, the local chess club President and also from the “*Glossary of Commonly used Terms in Chess*” from Wikipedia (Wikipedia contributors, 2020b)

To illustrate this work, a full example is given for a particular task displayed in Figure 7.3. This is a N-Check-Mate task, Black to play, with $N = 3$ (i.e. Black can win in 3 moves from this state of the game). This task involves a diverse set of concepts: at first sight, the situation appears unfavorable for Black, as they are outnumbered, their key pieces are poorly developed, the middle of the board is controlled by White and two key pieces of Black (Knight on E4 and Queen on G6) are directly threatened. However, players with enough expertise would notice that the White King can not move for two reasons: the protection of his allied pieces also blocked him, and the Black Bishop in C5 controls the square G1. Thus White King is vulnerable. A combination of attacks from Black Knight, Queen, and Bishop can lead to White King being checkmated. This sequence of moves is:

1. First move:
 - (a) Black Knight captures White Pawn in G3, checking White King;
 - (b) White is forced to capture Black Knight with White Pawn in H2.
2. Second move:
 - (a) Black Queen moves from F6 to H6, checking White King;
 - (b) White is forced to shield his King by moving his Bishop from G2 to H3.
3. Third move:
 - (a) Black Queen captures White Bishop in H3, check-mating White King.

The specific term "force to" for White means that this is the only legal move possible to play. Relevant relations that are involved in this sequence of moves are listed in Table 7.2 (page 114) and some relevant chunks are illustrated in Figures 7.5 (all of them are presented in Appendix B on page 146). Chunk information is presented using the Frame Theory from Minsky (see Section 6.1.1). Each chunk's image represents his spatial extent (as described in Section 6.2.1).

In total, over the 7 tasks elaborated for the second experiment, 93 chunks have been listed and detailed. This resulting chess ontology database will help us to understand players' fixations better, as well as identify which subset of chunks are considered by players before taking their final decision. We can also expect that each player has its chunk collection (acquire with practice), which does not precisely match our chunk definition.

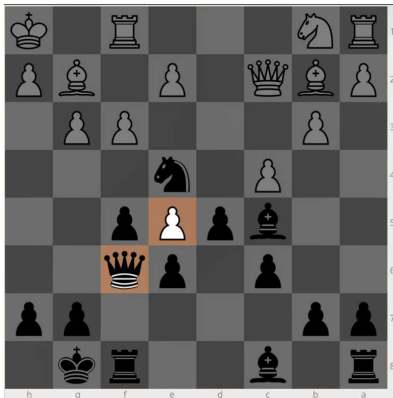


Figure 7.3: N-Check-Mate Task with N=3. Black to Play.

7.3.5 The Experimental Protocol

This second protocol is based on the one designed for our first experiment (see Section 5.3), with a few changes. Subjects were initially asked to play two easy practice games to become familiar with the equipment. We then performed an eye-tracking calibration before starting to record their eye-gaze, emotional state, and body reactions. The tasks are presented in this particular order:

1. Mate in 1. Black to Play.
2. Mate in 3. White to Play.
3. Mate in 3. Black to Play (with unbalanced situation).
4. Hopeless Situation. Black to Play.
5. Mate in 5 – White to Play (with unbalanced situation).
6. Hopeless Situation. Black to Play.
7. Hopeless Situation. White to Play.

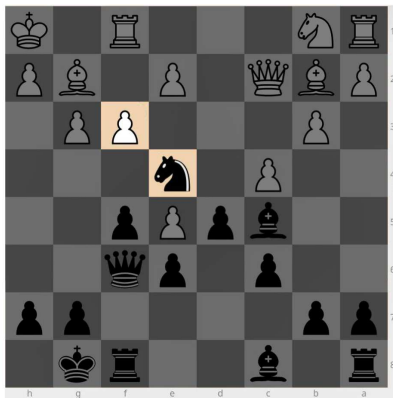


Chunk 1: White Pawn E5 Threatens Black Queen F6

```
(Chunk ChessChunk (C1)
  (Entities (WP5*), (BQ*))
  (Color (White))
  (Relation (Threatens))
  (Spatial-Extent (E5, F6)))
```

Entity **WP5** is the chess piece White Pawn E5.
Entity **BQ** is the chess piece Black Queen.

(a) Chunk 1: White Pawn E5 Threatens Black Queen F6

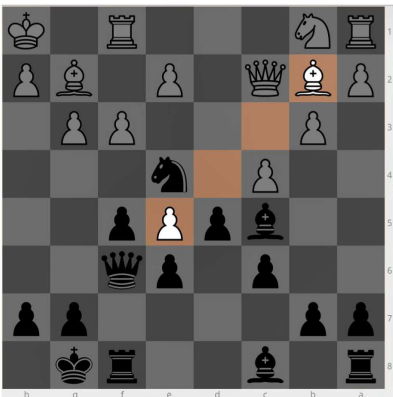


Chunk 2: White Pawn F3 Threatens Black Knight E4

```
(Chunk ChessChunk (C2)
  (Entities (WP3*), (BK1*))
  (Color (White))
  (Relation (Threatens))
  (Spatial-Extent (F3, E4)))
```

Entity **WP3** is chess piece White Pawn in F3.
Entity **BK1** is chess piece Black Knight in E4.

(b) Chunk 2: White Pawn F3 Threatens Black Knight E4



Chunk 3: White Bishop B2 covers White Pawn E5

```
(Chunk ChessChunk (C3)
  (Entities (WB2*), (WP5*))
  (Color (White))
  (Relation (Protects))
  (Spatial-Extent (B2, C3, D4, E5)))
```

Entity **WB2** is chess piece White Bishop in B2.
Entity **WP5** is chess piece White Pawn in E5.

(c) Chunk 3: White Bishop B2 covers White Pawn E5

Figure 7.5: Relevant chunks of N-Check-Mate ($N = 3$) Task presented using Frames.

Name	Subject Type	Object Type	Definition	Arity
Controls / Be Controlled	Piece	Square ; List of Squares	A piece controls a square or a list of squares, threatening all possible opponent pieces that enter this area.	2 or more
Threats / Be Threatened	Piece	Piece ; List of Pieces	A piece can capture another opponent piece in one move.	2 or more
Protects / Be Protected	Piece	Piece	A piece A protects an allied, piece B, meaning that if B is captured by an opponent piece C, A can directly threaten C.	2 or more
Screens / Be Screened	Piece	Piece	A piece screens an allied piece from a direct threat by being in front of it.	2
Pins / Be pinned	Piece	List of Pieces	When a piece is attacked and can legally move out of the line of attack, but such a move would expose a more valuable piece (or an unprotected piece) to capture.	3
Traps / Trapped	Piece	Piece	A piece that is unable to move for several reasons: its available moves are impossible due to the presence of others pieces, or available squares are controlled by opponent pieces.	1 or more
Sacrifice / Is sacrificed	Piece	Piece	A move or capture that voluntarily gives up material in return for an advantage such as space, development, an attack or saving a more valuable piece.	1

Table 7.2: List of Relevant Relations identified for the N-Chess-Task displayed in Figure 7.3

In contrast with the first experiment, no indication of what is at stake in the task is given. The player may play offensively or defensively depending on his playing style of play and on his understanding of the task. On completion of each task, subjects were asked to explain their understanding of the board situation, and the reason for their moves (RTE). We specifically asked them to identify opportunities, threats, and possible moves that were considered, even those excluded because they are seen as weak moves. The exact instructions for RTE given before the start of the experiment¹ are:

“At the end of each problem, I will ask you to explain your reasoning as well as possible. That is, name the pieces, piece configurations, positions, threats or opportunities you considered before making your final decision; Try to put yourself in the position where you want to explain your reasoning to a beginner or a child. Be as complete as possible: if you have considered moves that you then gave up, it is important to mention them. If you thought an opponent’s piece was threatening for a moment, before you discarded it, it is also important to mention it. If you see that an opponent’s piece may be dangerous, explain how, by which moves it may be threatening. State clearly the name of the piece and the important positions. For example, the Queen in C6 threatens the Rook in G2. All the steps you went through before making your first move. Any things you noticed after your first move should not be mentioned.”

If the participant is not very talkative (less than 30 seconds of speech) and, based on the moves he has made, the experimenter may intervene by asking one or more questions to encourage the participant to add information.

¹The experiments are carried out in French, the verbal protocols presented here are therefore translated.

These questions are, in this particular order:

1. Have you seen more pieces, positions, or configurations that you consider important before playing?
2. In what order have you identified the characteristic elements of the problem?
3. Do you want to add something before moving to the next task?

In any case, the third question is always asked before moving to the next task. If the participant exceeds 2 minutes of explanations, question 1 (or 2 if 1 is already asked) is not asked.

7.3.6 Metrics

The same metrics from the first experiment (see Section 5.3), along with the new modality of speech, are used for this new experiment. These are listed in Table 7.3, and what differs from the first experience is detailed in the following paragraphs.

Modality	Extracted Data	Computed Features	Computation Software
Body	Body keypoints	Body Agitation Body Volume Self-Touches	Openpose Homemade scripts
Emotion	AUs activation	Valence Arousal	OpenFace Homemade scripts
Gaze	Gaze and Fixation data	Scan Path Fixations Duration Fixations Order	EyeWorks
Speech	Retrospective Task Explanation (RTE)	Verbatim	Homemade scripts

Table 7.3: Recorded Metrics

Retrospective Task Explanation.

Verbal explanations are recorded during the entire experiment using the Kinect’s microphone before being transcribed in a text by hand. Afterward, the text is encoded in verbatim, and relevant concepts are extracted. Finally, these concepts are matched with our ontology database.

Emotion Analysis Software.

For this experiment, OpenFace (described in Section 5.2.3.1) has been used to extract emotional features: Action Units Activation. From these activations, valence and arousal values have been computed.

Eye-Tracking Device and Software.

The Fovio eye-tracking bar has been mounted vertically on the left side on the Touch-Screen to avoid being disturbed by hand movements of players. The software EyeWorks (describe in Section 5.2.3.2) has been used to create AOIs and compute gaze features: scant path and fixation information.

7.4 Results

We present in this section our exploratory results on a single task: a N-Check-Mate task with N=3. This task has already been presented in the previous Section 7.3.4 (page 111). Its main components are recalled here: Black to play, they look in an unbalanced situation as they are outnumbered with a weak pieces' development, two key pieces of Black are under threatened (Knight on E4 and Queen on G6). However, Black can still win in 3 moves which are:

1. First move:
 - (a) Black Knight captures White Pawn in G3, checking White King;
 - (b) White is forced to capture Black Knight with White Pawn in H2.
2. Second move:
 - (a) Black Queen moves from F6 to H6, checking White King;
 - (b) White is forced to shield his King by moving his Bishop from G2 to H3.
3. Third move:
 - (a) Black Queen captures White Bishop in H3, check-mating White King.

The main and relevant concepts are listed in Figure 7.5.

Due to the unbalanced position of Black, conservative players will be inclined on defensive moves such as protecting their threatened pieces and giving priority to the Queen's safety. The pieces that should attract the most attention for defensive players are C1, C2, C3, C4, presented Figures B.1a, B.1b, B.2a and B.2b respectively (page 146). These chunks involved pieces such as Black Queen in F6, Black Knight in E4, White Pawn in E5, White Pawn in F3, White Queen in C2, and White Bishop in B2. For experimented players, they will identify weakness in White's defenses such as White King being trapped by its pieces (White Pawn H2 and White Bishop G2) and because of the Black Bishop in C5 controlling the square G1. Attacking players would focus on the following chunks: C5, C6, C7, C8, C9, C10, C11 presented Figure B.2c, B.2d, B.2e, B.2f, B.2g, B.2h and B.2i respectively (page 147); involving the following pieces: Black Queen in F6, Black Knight in E4, Black Bishop C5, White King H1 and the following key squares H6, G3, H3 and G1.

We present in this section the analysis for three participants of different levels (novice, intermediate and expert). These players are interesting to consider because of the significant differences in their verbal reports and observations. For each player, we present the recording of their arousal when they observed and reasoned about the task in question. From these Arousal values, we target our Eye-Gaze analysis around the detected emotional peaks (which is delimited by red dashed-lines on every emotional plot). For this, we select the scan path, and the fixations performed 5 seconds before and after each emotional peak. Knowing the stakes and the chunks necessary to solve the task, we give a subjective analysis of these scan paths by pointing out the possible chunks observed by the player. Then, we apply the probabilistic formulas (developed Section 6.2.1) on the player's fixations, always centered on the emotional peaks. To validate the results (subjective observations and formulas), we finally give the verbal protocol of the player highlighting the chunks put forward by the player. We can thereby check whether the emotional peaks are indeed a good indicator to analyze the reasoning of the players.

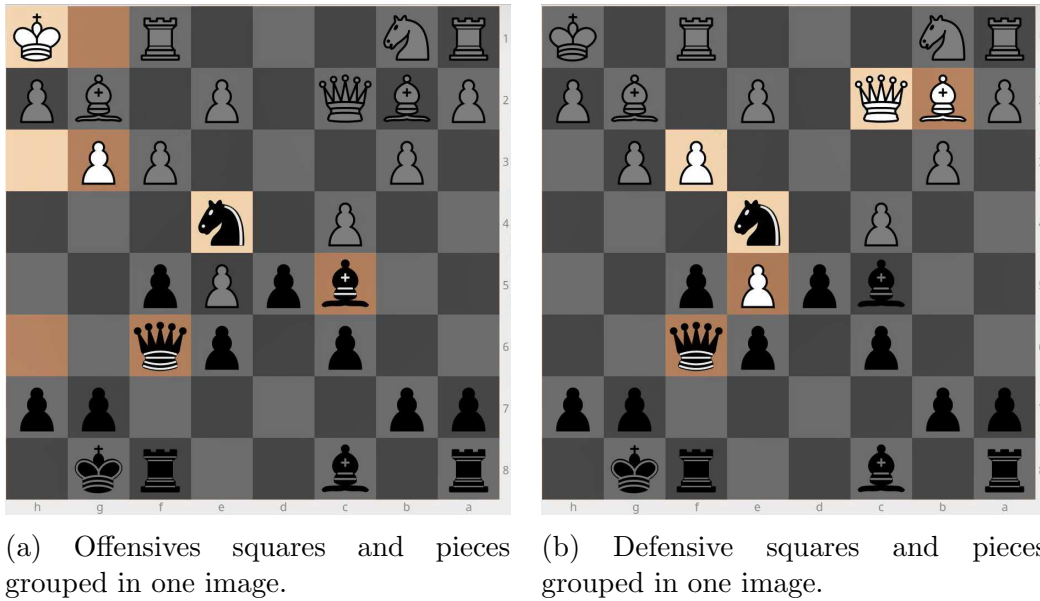


Figure 7.6: Offensive (left) and Defensive (right) Chunks.

7.4.1 An Expert Player

The participant identified the stake of the situation correctly and successfully solve it in 5 seconds. His explanation contains all offensive chunks that needed to be considered; no other chunk is mentioned.

Expert player's arousal is presented in Figure 7.7, between the moment the task is presented and the moment the player makes his decision ($t = 5s$), the average arousal value around 0.3 with a peak at 0.7 between the seconds 3 and 4. From the speed with which the player made his decision, we can easily conclude that the observation and reasoning phase was carried out between the seconds 0 and 5. These phases are delimited in Figure 7.7 by the two red dashed lines. Figure 7.8 displays the scan paths of player Q12 during his reflection phase ($t = 0s$ to $t = 5s$).

A first observation of the expert's scan paths reveals that the player only paid attention in the upper left part of the chessboard. A particular attention is paid on the Black Knight E4 and on the adjacent squares (containing both Black and White Pawns). On two occasions, at the end of the red scan path and the beginning of the blue scan path (thus one after the other), the player moves his attention from the upper left corner of the chessboard to F5 and G5, before going up quickly. This kind of behavior suggests that the player recovers the information from the bottom left corner of the chessboard with his parafoveal vision, this information contains the positions of Black Pawns and especially of the essential Black Queen in F6. After observing the bottom of the chessboard, in both cases, the scan path goes up towards the defensive position of the White King, suggesting that the player may consider offensively moving his Queen. Moreover, the diagonal of the blue scan path starting at F2 and ending at D5 corresponds to the diagonal covered by the Black Bishop at C5. The end of the blue scan path and the yellow scan path seems to focus on the Black Queen and its surroundings (threatening White Pawn and Black Pawns) before moving up to the upper left corner. Finally, the green scan path goes up to F2, one of the two squares accessible by the Black Knight to attack the opponent King. This same fixation, in F2, can also retrieve the information from the G1 square (covered by the diagonal of the Black Bishop C5) but also to the White Rook in F1 and the White Bishop in G2. The scan path ends with fixations around the Black Knight in E4 before ending on the G3 square, which is the second square accessible by the Knight to attack the opponent's King.

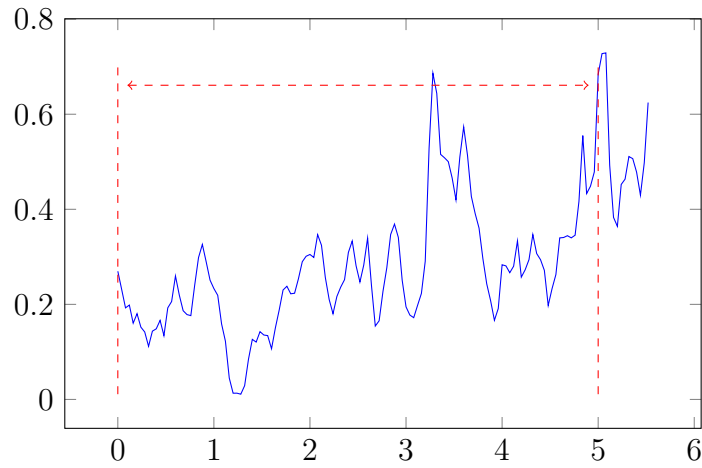


Figure 7.7: Arousal of an expert player (Q12). The observation phase extends from seconds 0 to 5. Figure 7.8 presents the scan path within the time frame defined by the two red dashed-lines.

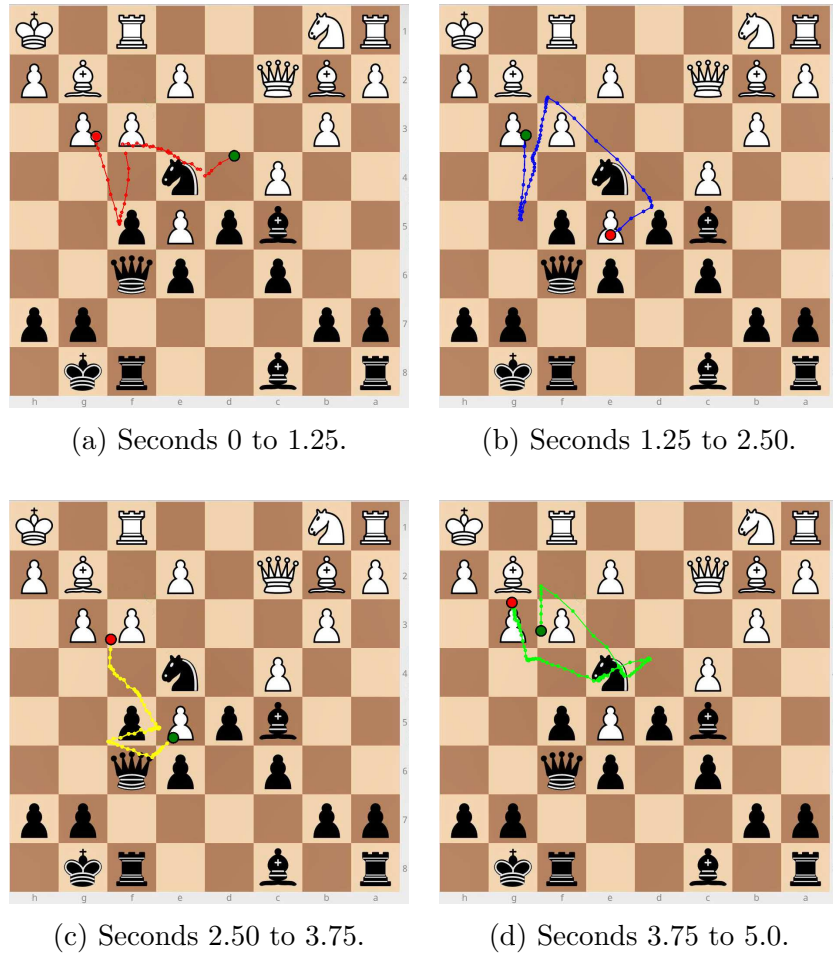


Figure 7.8: Time Windowed Scan Paths (1.25 seconds each) of the Expert Player from Seconds 0 to 5. The green and red points indicate the starting and the ending edges of the scan path, respectively.

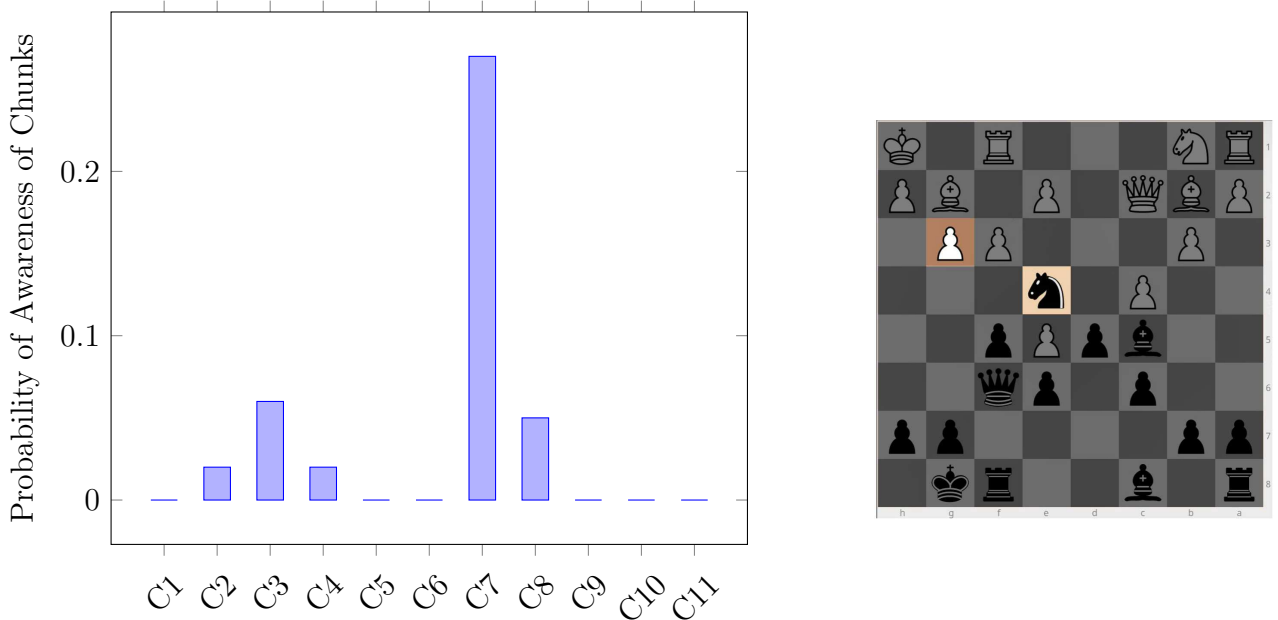


Figure 7.9: Left: Awareness of Relevant Chunks by Participant Q12 (expert) from his Fixations between $t=0s$ and $t=5s$. Right: Chunk C2.

Verbal Report

The subject justified his moves with a complete and concise explanation:

*“Ok I see that there is a mate in 3 here. It is kind of a pattern because again the **King is not able to move at all** [C9]. So I check if there is any check. There is 3 pieces here, Queen F6, Knight at E4 and Bishop at C5. So the **Bishop already controls the only available square of the White King** [C6], so there is two pieces. So first check with the **Knight** [C7] then the **Queen** [C10, C12].”*

All relevant and offensives chunks were mentioned. This implies that he successfully recognized these chunks from LTM and combined them to generate winning moves in less than 5 seconds.

Interpretation

Player Q12 is the participant who solved this problem the fastest, in 5 seconds. Observing the arousal response, a slight drop is visible between the seconds 1 and 2. This interval corresponds to the red and blue scan paths, during which the Black Knight is observed, and the imminent threat to it may be perceived by the player, causing a drop in his arousal. The peak at 0.7 of the arousal value is interesting considering the scan paths that preceded it (red, blue, beginning of yellow). During these scan paths, the relevant pieces for the task resolution are almost all fixed: the Black Knight E4 [C7], the square where to move the Knight G3 [C7], the Black Queen F6 [C10, C12], the diagonal of the Black Bishop C5 to G1 [C6] and the defensive position of the White King located in the upper left corner [C9].

Applying the formula defined in equation 6.7 to compute the probability of awareness of chunks using evidence from observations, the Figure 7.9 displays the most probable chunks considered by participant Q12 from its fixations. Chunks C7, which represents the first move to be played in the winning sequence, has the highest probabilities to be considered by the expert.

These chunks are all mentioned very clearly in the player's verbal report. This confirms that the fixations did indeed extract (among other things) these chunks, which led the player to understand how to solve the problem. This peak at 0.7 in arousal can, therefore, be interpreted as an indicator of the identification of a significant opportunity in the problem to be solved.

7.4.2 An Intermediate Player

The participant grasped a good understanding of Black's situation in 20 seconds. However, he could not find any good solution to either attack his opponent nor save his pieces. After having to sacrifice his Black Knight from E4 to F2 (threatening White King to save his Black Queen), the subject admitted that he was not satisfied by his final choice.

The value of player Q2's arousal is presented in Figure 7.10. The peaks at $t = 1s$ and the one $t = 20s$ correspond to excitation generated at the discovery of the game and the beginning of the game phase, respectively. Besides that, the value is around 0.43 between the second 0 and 12.50, where a decreasing tendency disturbance can be observed. We observe this perturbation in more detail by presenting the scan paths of the player between the seconds 10 and 20 with the Figure 7.11. Each scan path represents 2 seconds. The first scan path, in red, is composed of fixations and saccades operating back and forth between the diagonal D4 and G1. These two-way scans can extract information about the Black Knight in E4, the threatening White Pawn in F3, the defensive White Rook in F1, the White Bishop in G2, the square F2 which is an offensive position for the Black Knight E4 and finally the diagonal covered by the Black Bishop C5 to G1. The second scan path, blue, again makes the same back and forth movement as the first seconds, but also observes one end of the upper right corner of the chessboard, containing the White Queen C2 threatening the Black Knight E4. The scan path comes back to Black Knight E4 after noticing the White Queen C2. The yellow scan path becomes narrower around the F2 square and the Black Knight E4, suggesting that the player is thinking of moving the Knight to this position. Then the scan path goes down to the position of the Black Queen F6 and the White Pawn E5 threatening it. The green scan path again considers the attacking position F2 of Black Knight E4 before checking the cover of Black Bishop C5. A novelty is the attention paid to Black Bishop C8 before moving up to the endangered Black Queen F6, suggesting that the player is looking for new possibilities. Finally, the purple scan path reveals that the player's attention is clearly focused on the Black Queen F6 and the Black Knight E4, and the choice seems difficult because of the many backs and forth moves between these two pieces.

Verbal Report

Subject's verbal analysis mainly focused on the two pieces under attack, the Queen and the Knight, and the possible ways to save efficiently one of them. No other relevant chunks are mentioned.

*“So the **Black Queen F6** is threatened by the **E5-Pawn [C1]** and in addition, the **E4-Knight** is threatened by the **F3-Pawn [C2]**. So I have two threatened pieces at the same time, so it's not going to go well and in principle I'm going to lose one of them. So I thought that if I uses the **Knight to check [C8]** him (Note: the participant refers to moving Black Knight E4 to F2), he would have to answer and that would allow me to move the Queen. So I checked with the **Knight [C8]**, he takes back with the Rook, and now, I have to move my Queen but I have no way to protect my Knight so that's it.”*

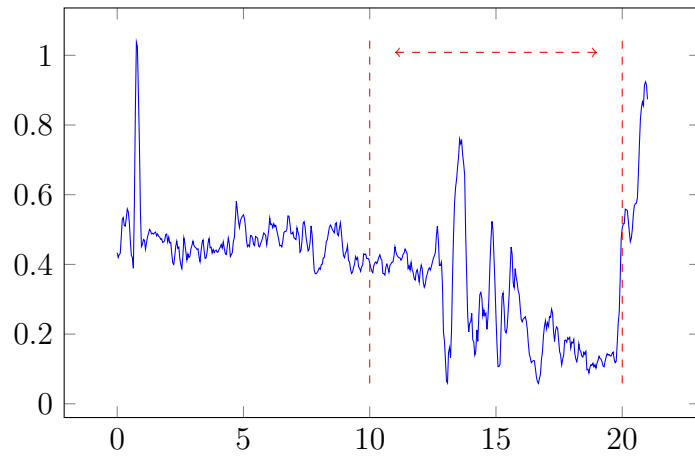


Figure 7.10: Arousal of an Intermediate Player (Q2). Figure 7.11 presents the scan path within the time frame defined by the two red dashed-lines.

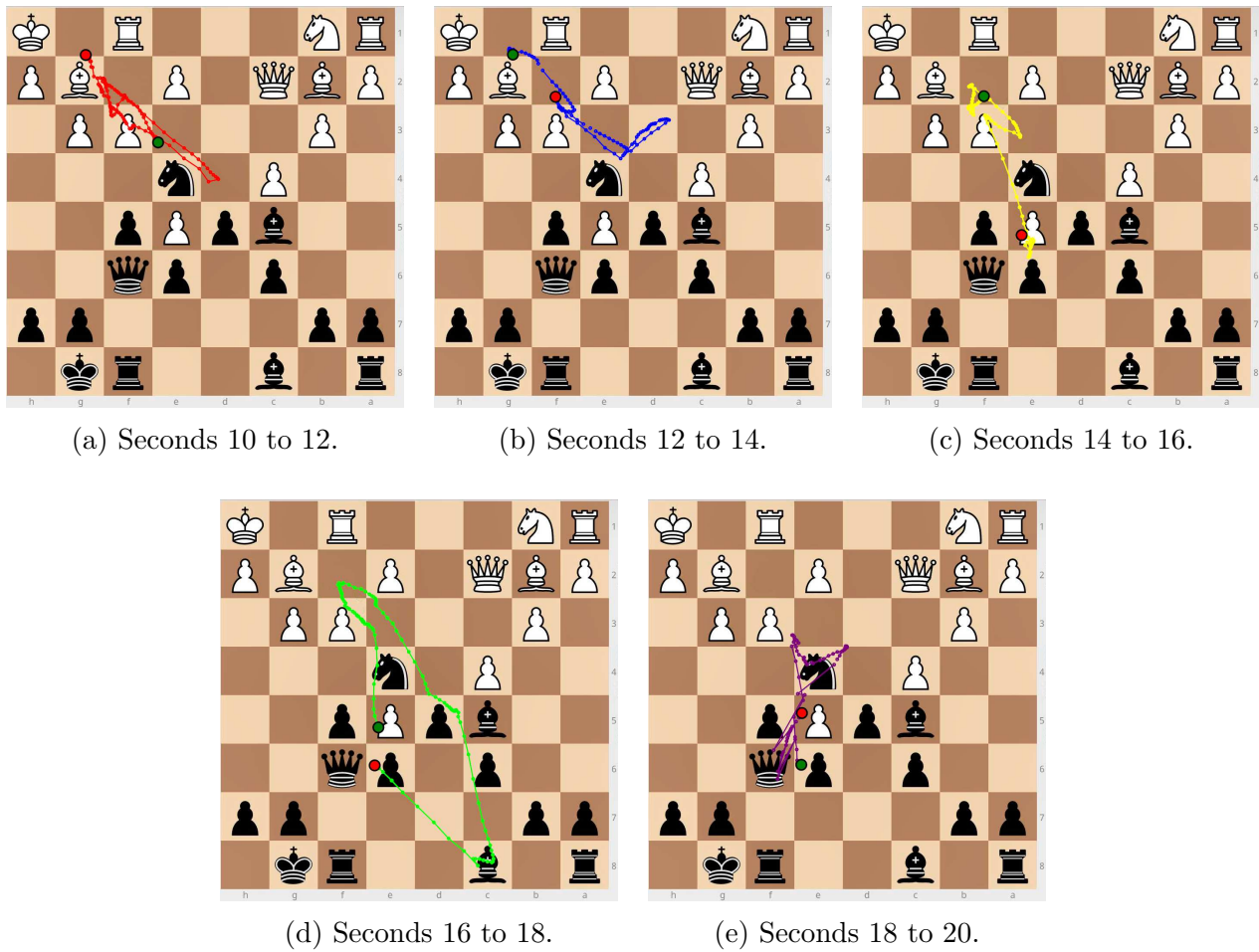


Figure 7.11: Time Windowed Scan Paths (2 seconds each) of the Intermediate Player from Seconds $t=10s$ to $t=20s$. The green and red points indicate the starting and the ending edges of the scan path, respectively.

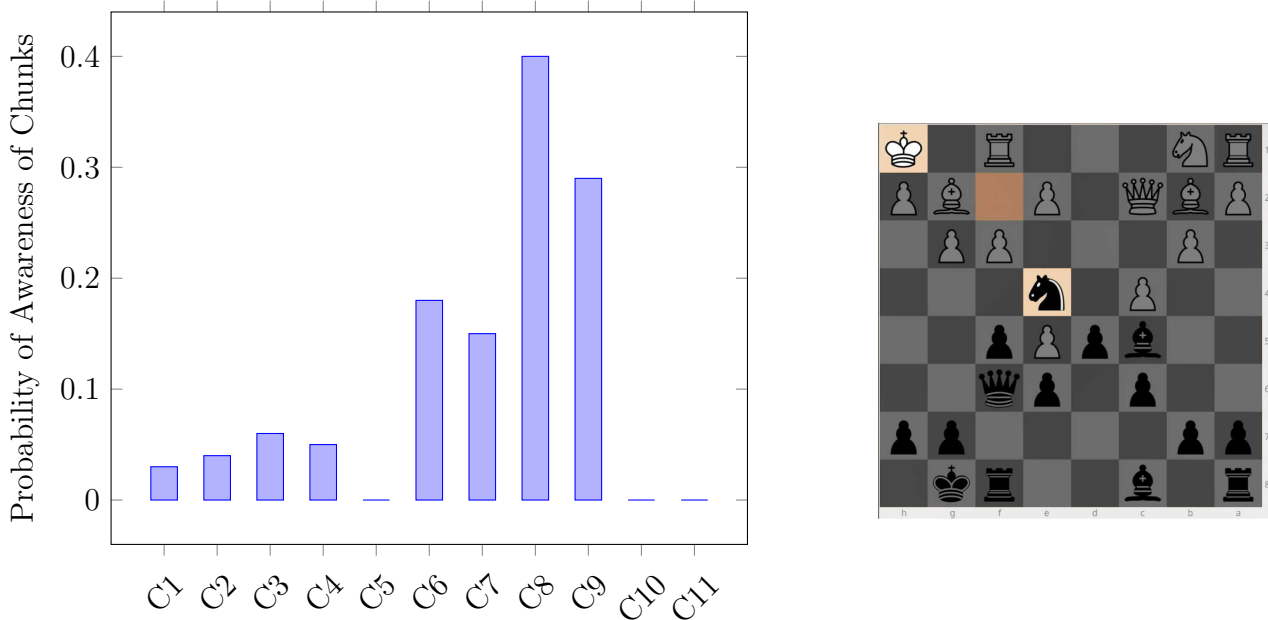


Figure 7.12: Left: Awareness of Relevant Chunks by Participant Q2 (Intermediate) from his Fixations between $t=10s$ and $t=20s$. Right: Chunk C8.

Interpretation

The intermediate player pondered for 20 seconds before making the decision to sacrifice his Black Knight E4 in F2 to gain time and save his Black Queen F3. The emotional arousal value of the player shows a disturbance from $t = 10s$, which decreases the average value until $t = 20s$, the time when the player makes his decision and plays. By observing the player's visual attention, we can see that between $t = 10s$ and $t = 15s$, most of the attention is focused around the Black Knight E4 and the surrounding squares of F2 ² ([C8]). The second part of the scan paths (between $t = 15s$ and $t = 20s$), focuses on the threatened position of the Queen ([C1, C3]) and possible support (observation of the Black Bishop in C8). However, the last 2 seconds of the scan path shows a difficult choice to save his Queen or his Knight from the imminent attack ([C1, C2]).

Figure 7.12 displays the probability of awareness of chunks from the participant's fixations. Chunks C6, C7, C8 and C9 are the most probable according to the formulas. The player's verbal report confirms that all his attention has been focused on his two threatened pieces: the Black Knight and the Black Queen. The player admits that he realized the ineluctable loss of one of them and explains his solution to making an attack in F2 ([C8]) with his Knight to gain time and save his Black Queen. However, the participant did not mention chunks C6, C7, and C9 despite their high probabilities of awareness. This is explained by the overlapping of Spatial Extent of chunks C6, C9 and C8 which is the most considered (see Images B.2d, B.2f and B.2g on page 148). This awareness in the difficult position of the player can explain the disturbance of the player's arousal, and his excitement decreases until the final decision is taken.

²Moving the Black Knight E4 to F2 corresponds to the chunk [C8], this attack puts the opponent's King in check, forcing the opponent to defend his King by capturing the Knight with his White Rook in F1. However, placing the White Rook in F2 places it under the threat of the Black Bishop C5. This attack by the Knight, even if it puts the opponent's King in check, does not give the Black player any advantage because if he wants to save his Queen once the Knight's attack has been launched, he must renounce capturing the White Rook posted in F2 once it has seized the Knight.

7.4.3 A Novice Player

The novice player was unable to solve the task after almost 80 seconds of effort. He focused his time on saving his Queen and was unable to consider any opportunity offered by an offensive move. He finally moved his Black Queen to safety from F6 to G5. Sacrificing the Knight to save his Queen was evident, and he did not consider doing something useful with the Knight before losing it.

The player's arousal, Figure 7.13, during these 80 seconds is close to 0.2 but also has many peaks. Some peaks, of average height, are in 0.4 and 0.8. Two other peaks ($t = 41s$ and $t = 64s$) reach 1.4, the scanpaths corresponding to these peaks are respectively displayed Figures 7.14 and 7.15.

Between $t = 36s$ and $t = 41s$, the attention is focused on the two White pieces: White Queen C2 threatening Black Knight E4 and White Bishop B2 protecting the White Pawn at E5 (and by the same occasion, preventing the Black Queen from capturing the White Pawn E5 to free herself from his threat). Several backs and forth moves are made between the White Pawn E5, who threatens the Black Queen F6, and the White Bishop B2, which offers a cover to the White Pawn. Then visual attention moves to the square F2 and the White Rook in F1. The player's gaze also considers the Black Knight E4 before returning to square F2, suggesting that he may consider the offensive possibility of attacking with the Black Knight in F2. The scan paths concerning the peak $t = 64s$ are, on the other hand, very focused in precise locations. The player makes many long fixations on the squares G3 (White Pawn), G2 (White Bishop), F3 (White Pawn), and E4 (Black Knight). These long fixations suggest a possible strategy of the player concerning these squares. The possibility to attack with the Black Knight E4 in G3 or F2 is difficult to infer from these scan paths, which do not present successive fixations between these squares.

Verbal Report

The novice players' explanations focus on ways to secure the Black Queen. For this player, the Knight was already lost, and moving it was useless. Thus effort was focussed on moving the Queen. He admitted that a better solution may have existed, but could not find it.

*“So here I saw that the **Queen was being attacked by the Pawn** [C1], it was close to the middle so it was easy to see. Then I tried to scan a lot of things. I thought of looking at how to **eliminate this one** (note: White Pawn E5)[C1] without having to leave from there (note: F6). I couldn't think of any other solution than to remove the **Queen** [C1]. I don't know if there was one. I thought that, if I was going to move it (Note: the Black Queen), I might as well do something that would allow me to get closer and clean the Pawns that were over there (Note: White Pawns in G3 and F3), that's why I did it diagonally (Note: move the Black Queen diagonally from F6 to G5.). So I saw at first that I had to **sacrifice one of them** (note: pointing at Black Queen and Black Knight) [C1, C2], but maybe that's where I went wrong and maybe I could have saved them both.”*

Interpretation

In 80 seconds of observation and reasoning, the novice player has scanned the chessboard several times and considered several strategies. The emotional arousal graph suggests several opportunities discovered by the player. When we look at the scan paths of the two most significant peaks, we discover the primary concern of the novice player: the imminent attack on his Queen [C1]. This concern is reflected in the numerous backs and forth between the White Pawn in E5 threatening the

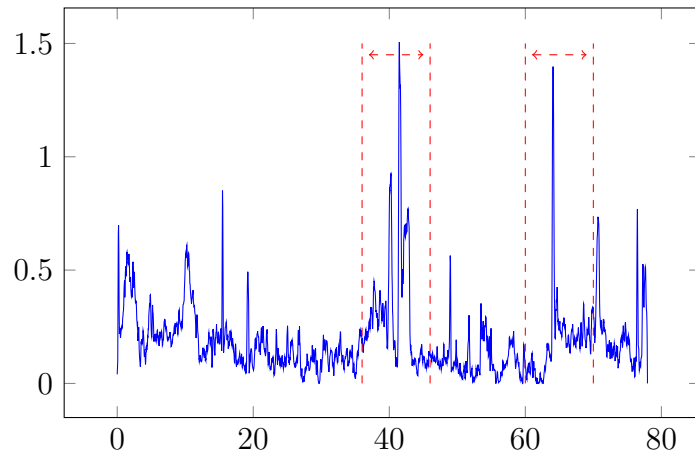


Figure 7.13: Arousal of a Novice Player (Q19). Figures 7.14 and 7.15 present the scan paths within the time frames defined by the red dashed-lines.

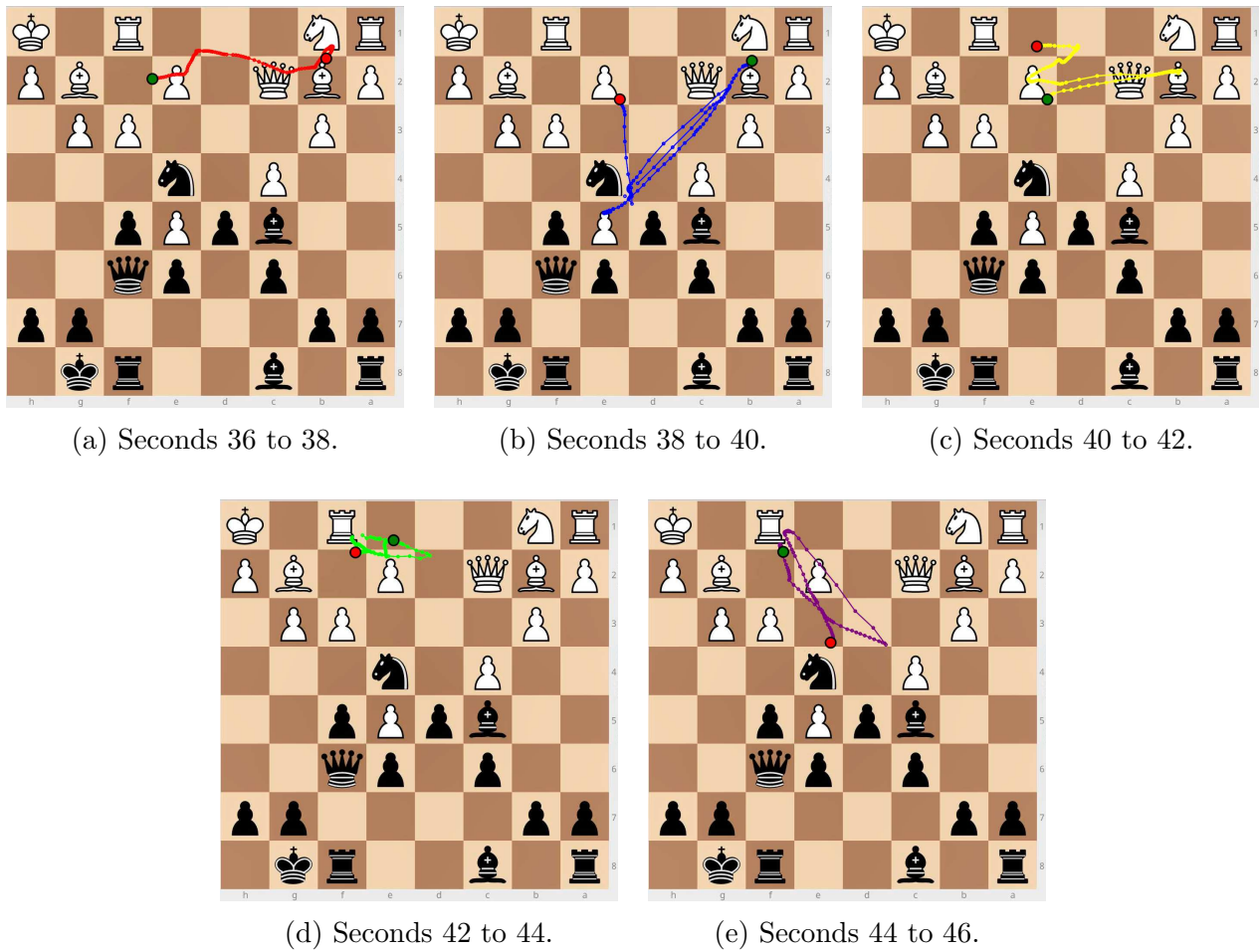
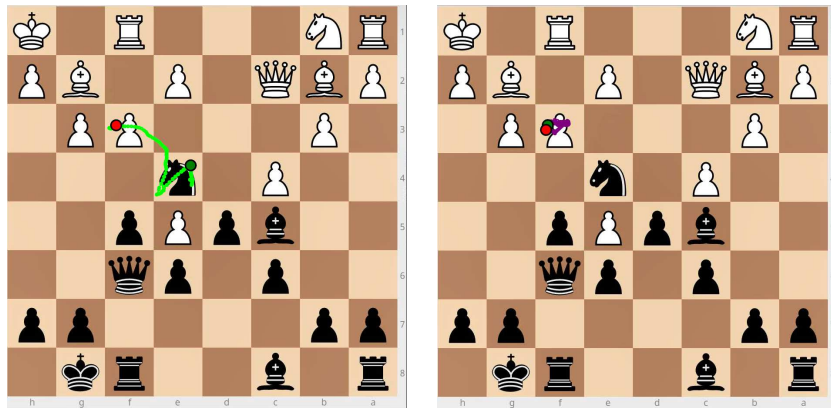


Figure 7.14: Time Windowed Scan Paths (2 seconds each) of the Novice Player from Seconds $t=36s$ to $t=46s$. The green and red points indicate the starting and the ending edges of the scan path, respectively.



(a) Seconds 60 to 62. (b) Seconds 62 to 64. (c) Seconds 64 to 66.



(d) Seconds 66 to 68. (e) Seconds 68 to 70.

Figure 7.15: Time windowed scan paths (2 seconds each) of participant Q19 from seconds 60 to 70. The green and red points indicate the starting and the ending edges of the scan path, respectively.

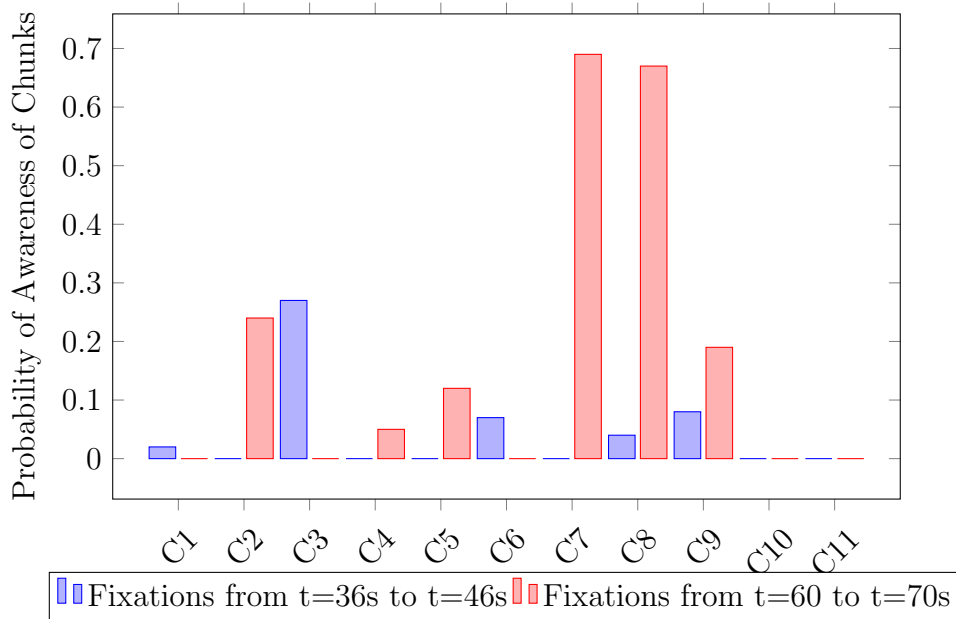


Figure 7.16: Awareness of Relevant Chunks by Participant Q19 (Novice) from his Fixations.

Black Queen [C1] and the White Bishop B2 supporting the White Pawn [C3]. The probabilities computed to determine which chunks are considered depicted Figure 7.16, also suggest that the participant was first (blue barplot) focused on the threatening White Bishop and White Pawn [C3]. Furthermore, this concern is confirmed in the player’s verbal report, who admits to trying to save his Queen [C1] and let his Knight be captured [C2]. The follow-up of the scan path of Figure 7.14 seems to be interested in action possibilities in the upper left corner of the chessboard, where only the Black Knight E4 [C7, C8] and the Bishop C5 can access [C9]. For this sequence, the probabilities (red barplot of Figure 7.16) show a high interest in these chunks. However, these strategies seem to have been abandoned because none of them are mentioned in the player’s verbal report. The strong arousal spike could, therefore, reflect the player’s strong concern about his inability to protect his Queen effectively, this concern can be inferred from the backs and forth fixations preceding the spike and the corresponding chunks in the player’s verbal report.

The second peak is more challenging to interpret. Indeed, long fixations reflects difficulty in extracting information or in making a decision (see Section 4.1.3). One strategy seems to be considered carefully by the participant. According to the position of the fixations, this strategy would concern White Pawns G3 and F3, White Bishop G2, Black Knight E4 (chunks [C7, C8, C9]). However, no clear relationship can be deduced from the scan paths, because no saccade seem to bind the pieces together (except the threat of White Pawn F3 on Black Knight E4 [C2]). By reading the verbal report, we understand the strategy considered by the player. He was concerned with the second move, the one he would play once he got his Queen to safety. However, we did not list the chunks involved in this situation. As we have been able to note in the verbal report and the final moves played, the player judged that the strategy considered before the emotional peak at $t = 64s$ was interesting to play. Indeed, the player states: *”I thought that, if I was going to move it (Note: the Black Queen), I might as well do something that would allow me to get closer and clean the Pawns that were over there (Note: White Pawns in G3 and F3), that is why I did it diagonally (Note: move the Black Queen diagonally from F6 to G5)”*.

In summary, for the two most intensive peaks of arousal, we have a decision concerning the pieces fixed by the player a few seconds before the beginning of the peak. We can infer, by these results, that these high-intensity peaks are correlated with decision making.

7.5 General Discussion

This second experience provided new insights into understanding the role of emotions in reasoning and decision making. In our first experience, we noticed that the observed emotions were almost neutral when solving trivial problems (requiring reactive play). In contrast, when faced with more and more challenging problems, the emotions varied very quickly, in the form of short peaks. To study this behavior, we introduced several new elements for our second experience:

- A Cognitive Model for representing the concepts and knowledge of the players (Frames) and representing the working memory using Situation Models;
- A Verbal Protocol (Retrospective Task Explanation) asking players to explain, after task completion, strategies they consider to take their final decision;
- New Tasks to elicit emotions: more complex, combining several different strategies (offensive or defensive) and presenting critical situations for the players;
- An ontology of 93 chess concepts that lists all relevant chunks for each task. This ontology lets us identify and code very quickly the concepts put forward by the participants during the verbal protocol.

As a result of this experience, we present three complete examples of players of different levels: an expert, an intermediate, and a novice. In our first experiment, we were blinded by the incomprehensible accumulation of player-generated scan paths. It was impossible to deduce a strategy with a scan path from a player who had been thinking for more than 30 seconds. This time, the emotional values of the players guided our analysis. For each player, we place side by side the emotional data of arousal and the scan paths of the players preceding and following by a few seconds the peaks we could observe in their emotional graph. For each analyzed peak, we were able to conclude that the pieces and configurations of pieces fixed a few seconds before the appearance of this peak was 1 - always mentioned in the verbal report of the player and 2 - always involved, directly or indirectly, in the final decision (moves played). Based on these results, we can, therefore, deduce the representation of the players' concepts present in their Working Memory at the precise moment of an emotional peak by analyzing the fixations produced a few seconds before the appearance of the peak. We can also conclude that the most intensive peaks reflect a decision making, either intermediate or final, which impacts the final decision. This last conclusion is validated by the content of the players' verbal reports.

However, we can only infer the content of the working memory and the representation of the players' knowledge, we can not for the moment deduce it systematically. No systematic procedure has yet been developed, and our results are based on a subgroup of participants. Our hypothesis proposing the correlation between emotion and decision making is confirmed by the study of players' arousal peaks and their scan path. Thanks to the verbal report, we can confirm our deductions made from the emotional and visual data, consolidating our conclusions.

Besides, no information on the body behavior of the players has been considered yet in our results. A possible next step would be to compare body agitation and self-touches around the emotional peak to observe if there is any apparent increase (or decrease) in body behavior at that moment.

One of the hypotheses proposed before the beginning of Experiment 2 was: Could the content of Working Memory of chess player be inferred using their Verbal Protocol? If it is true that the concepts put forward by the players in their verbal protocol must have been present in their working memory at a given moment, it is difficult to predict the exact content of working memory at a given moment t . Working Memory has a limited storage space, and the oldest or no longer relevant elements are ejected from this space to make room for new concepts considered by the player. However, the verbal protocol allows us to confirm that a chunk is perceived by the player and should be present for a moment in his working memory. To determine when this chunk is actually present in working memory, we must correlate the emotional peaks and the scan path as we did for the 3 players presented as examples.

Our cognitive model was partially used in this experiment. The Frames allowed us to code and identified the chunks put forward by the players in their verbal protocol. Formulas to extract information from evidence gave interesting results that are close to the verbal report of players. However, these formulas need to be improved as some naive assumptions were made, and more prior knowledge can be easily integrated. The next step would be to improve the information extracted from a fixation. The current state only consider the information of the fixated square without considering a potential use of parafoveal vision. A more reasonable approach would consider a gaussian centered on the fixation of a size of approximatively 3x3 squares grid. The gaussian size could also be proportional to the fixation duration. All squares and pieces capture inside the gaussian would be counted as perceived, with equally distributed energy or attributing more energy to the center. Nevertheless, we think that these formulas would allow a systematic approach to the analysis of the pieces and configurations of pieces considered using emotions and visual data. We have presented a preliminary version of what this cognitive model would yield

once fully developed, and the results are promising.

Finally, the development of our protocol also benefited greatly from the lessons learned from our first experience. Because the physical limitations were known, we had better control over our records, avoiding any loss of data. The use of an eye-tracker on the side of the screen enables us to achieve stable results with minimal data loss. However, we still can not prevent the player from moving around in his chair, sometimes moving away from the eye-tracker's field of action.

Chapter 8

Conclusion and Perspectives

8.1 Summary

Chapter 2 opens with a description of two players engaged in solving a chess problem. These players, a novice and an expert, presented two distinct behaviors: Nemo, the novice, looked in great difficulty to solve the problem, this nervousness was reflected by physiological variations (accelerating heart rate, agitation on the chair, repeated self-touches, activation of facial micro-expressions) and by numerous eye fixations (back and forth from one end of the chessboard to the other, repeated fixations on several pieces). Finally, he did not succeed in solving the problem. Elliott, the expert, was calmer, he presented few somatic signs and few eye fixations, he found the solution very quickly. We concluded these observations by asking three questions, the sections and chapters that followed them gave some clues to answer them:

1. How one can explain the somatic variations of players?

Chapter 2, 3 and 4 have given several answers to this question. First, in Chapter 2, the structure, and functioning of the human memory have been studied to highlight the mechanisms used by chess players. We have seen how knowledge and cognitive elements are stored in short-term memory and in long-term memory (Atkinson and Shiffrin, 1968; Broadbent et al., 1986; Cowan, 2016). Working Memory has been further detailed due to its importance in problem-solving. Its limited storage system forces players to perform the cognitive process “chunking,” which allows them to encode more meaningful information in the same slot in Working Memory (Miller, 1956). Hebb’s law allows us to understand how players can gather their learned skills needed very quickly in their working memory to solve a problem (Hebb, 1949). The intuitive moves played can be explained thanks to the “Dual-Process” theory: System 1 and System 2 (Kahneman, 2011). System 1 is fast and solves perceived problems by pattern matching with little control over the answers, unlike System 2, which can engage energy-intensive cognitive processes to solve problems. However, System 2 is lazy and tends to do everything to make System 1 do most of the work. The cognitive processes used by System 2 do not go unnoticed. This is what Damasio explained with the Somatic Markers theory by stating that conscious and unconscious cognitive processes are reflected by somatic (**e.g.** physiological) external signs (Damasio, 1996)

Secondly, Chapter 3 offers a non-exhaustive but rich review of the different definitions and models proposed to study human emotions in recent decades. Several definitions, starting with that of William James, are given (William and Harter, 1899). We learn that emotions are generated from complex systems and are not resulted from a single process but rather by complex communications between several parts of the body and the brain. Several models of emotion representation

are then presented. The most well-known are those based on the activation of micro muscles, proposed on the one hand by Ekman: Basic and discrete emotions (Ekman, 1957, 1977); and by Russell and Mehrabian who present emotions as continuous values representable on a 3D model: Pleasure-Arousal-Dominance (Russell and Mehrabian, 1977; Russell, 1980). However, these models are not the only ones considered and are even criticized (Barrett, 2011). It is essential, even crucial, to consider the context in which the emotions are generated as they can be interpreted differently (Barrett et al., 2011).

Neurobiological studies have also provided valuable information about emotions and the key roles of brain organs such as the hypothalamus (Papez, 1937). Emotions can be described as the result of a chemical reaction of several neurotransmitters: serotonin, dopamine, and nor-adrenaline. Research also shows that some elements of memory are more prominent than others, depending on the emotions displayed during an event. This time, amygdala activity may be partly responsible for improved memory when excited by emotional hormones (McGaugh, 2000). Besides, if an event is experienced again, the same parts of the brain will be reactivated and would again experience the same emotional feeling (Buchanan, 2007).

Latter, in Chapter 4, we have seen that significant information can be observed in the players' eye fixations. Various research studies have shown the role of fixations and scan paths in information retrieval and retrieval (Duchowski, 2007; Holmqvist et al., 2011). Besides, studies have shown that unconscious body behavior reveal information about a mental state overloaded by stress or by too much expensive cognitive work (Harrigan, 1985; Aigrain et al., 2016b).

By presenting the structure of memory, its role in reasoning, and then looking at the models of reasoning, emotions, and their somatic signs, we have given many elements to understand the somatic variations of players during problem-solving.

2. Are they comparable?

Chapter 4 was focused on observable manifestations of human reasoning. We first discussed the role of the gaze and its functioning with an anatomical description of the eye and then of the human visual system. Several eye-tracking methods are presented, including one in particular: Video-Based Combined Pupil and Corneal Reflection that we applied with a Remote Eye-Tracking Bar. Elements of cognitive processes can be extracted from the gaze: the fixations show the elements that draw the attention of the subjects during their reasoning, and the scan paths suggest the sequence of reasoning steps. Numerous studies have focused on the expert information that can be extracted from the players' fixations (Reingold et al., 2001; Reingold and Charness, 2005). Significant differences can be observed in players' fixations, depending on their level of play. A player's expertise also influences his parafoveal vision. Indeed, the useful angle with which a player can extract information is all the more important as his expertise in the game is essential.

In a second step, we have seen how, from off-the-shelf and cheap cameras, we can identify the activation of facial micro-expressions based on the taxonomy proposed by Ekman (Ekman, 1977; Friesen et al., 1983) and using image processing software (Technology, 2015; Baltrusaitis et al., 2018). This information can then be used to recover the emotional state of the observed subject by combining the perceived activations of subsets of micro muscles and representing them either in the model proposed by Ekman (basic emotions) (Ekman, 1957), or in Russell's PAD model (valence, arousal, dominance) (Russell and Mehrabian, 1977).

Finally, the last section of the chapter 4 describes some studies made on the interpretation of body movements. First of all, we describe stress by three different approaches: biological, phenomenological, and behavioral (Aigrain et al., 2016b). Several observable characteristics of stress are presented: behavioral and physiological features (**e.g.** Body features, Posture changes,

self-touching, etc.). Special attention has been given to self-touching, due to the ease of observing it for chess players and the emotional and cognitive information it can provide. These body movements are interpreted as an indicator of the subject's emotional state (Ekman and Friesen, 1969). Other studies define these movements as unconscious and are used to briefly distract the brain's attention to help it discharge the excessive cognitive load caused by a stressful external stimulus (Troisi, 2002). Finally, we insist on the fact that, despite the significant informative elements provided by self-touches, these body movements are the result of complex systems and should not be assigned only to negative emotional states (Harrigan, 1985).

Once these manifestations of human reasoning were identified, we developed a recording instrument to observe these physiological and eye-related signs in chess players. The chapter 5 first describes this instrument, composed of cheap and commercially available sensors: RGB camera, Touch-Screen, Kinect, Eye-Tracking Remote, USB LED. The software used to record and analyze the data are also detailed. After validating our experimental protocol, we organized a first experiment to observe the reasoning manifestations identified in chapter 4 and to study them. 23 subjects participated in this experiment; their emotional reactions, eye trajectories, and body behaviors were recorded while they were engaged in solving 13 specially designed chess problems of increasing difficulty. The data collected yielded very intriguing results. Expert and intermediate players showed significantly different features. We were able, through a simple machine learning algorithm (SVM), to identify with more than 90% accuracy if the observations came from an expert or intermediate player. However, the emotional results were surprising. The incessant emotional variations and observable punctual spikes were not in agreement with our initial experimental hypotheses. We, therefore, decided to delve into the role of emotions in reasoning.

3. Can we tell, from what we see, their awareness of the situation and comprehend the threats and opportunities?

Awareness of something is related to a particular state where knowledge and memories of that thing are directly accessible. The Chapter 6 proposes a framework to model the understanding and awareness of chess players by the observations described in previous chapters. We proposed to model cognitive concepts using Minsky's theory of Frames (Minsky, 1974), which aims to unify Artificial Intelligence and Psychology concepts to better model human reasoning. Frames are data-structures used to describe specific entities, relations, or chunks (hierarchy of several sub-frames). We have seen how to represent simple chess concepts and other more complex ones, like chunks or complete chess situations.

Frames can be used to model the content and manipulation of concepts in Working Memory. For this we have proposed a Working Memory model, based on the literature (Hebb, 1961; Atkinson and Shiffrin, 1968; Miller, 1956; Cowan, 2016). This model represents the limited buffer, its interactions with the other components of the memory (episodic memory, procedural memory, concept knowledge) and the interactions with the outside world: sensory (vision, auditory, tactile, olfactory, gustatory, etc.) and manipulable (speech, manipulation, mobility, emotion display).

The observations described in Chapter 4 and from our first experience (Chapter 5) provide evidence on the subject's situational awareness. We proposed probabilistic formulas in Chapter 6 to compute pieces of evidence of awareness from the recorded observations. Using these formulas, eye-fixation and emotion can be directly interpreted by updating the subject's working memory at a given time. The state of his working memory and emotional observations would allow us to link the role of emotion in reasoning. Our hypothesis is that the subject uses the evoked emotions to select from the many possible situations for reasoning about moves during orientation and exploration.

To validate and deepen our model, a second experiment is described in Chapter 7. In the perspective to improve the information collected on chess players, we added in our experimental protocol a phase of self-report to define which chunks were used by the players to make their choice. We first present a study on the different methods of self-report to select the one that best fits our experimental environment. We choose a Retrospective Task Explanation (RTE) method, which consists of asking players to explain, thoroughly and concisely, the chunks and strategies considered to make their final decision. In order to identify the chunks identified by the players during their verbal protocol, we first had to list all the chunks characteristic of the tasks we performed. An ontology of 93 chunks has been set up for chess problems designed for our experiments. 23 participants took part in this experiment. On 3 players of different levels, novice, intermediate, and expert, we showed the relevance of the role of emotional variations in their reasoning. We were able to relate emotional peaks to eye trajectories to emphasize the crucial moments during which important decisions are made. These first results confirm that correlations can be made between emotions and choice of strategies. However, it remains difficult to establish a link between emotions and the other modalities (eye-gaze and body) because the timings are different, and a chunk can remain in Working Memory for few seconds to several minutes according to its utility. We remain convinced that we are on the right track and that adjustments can be made to the experiment to improve our correlations and results.

8.2 Contributions

An instrument to observe human engaged in problem-solving

We have set up a multi-modal instrument that effectively and robustly captures the reasoning manifestations of chess players. This equipment is based on off-the-shelf equipment. With the exception of the eye track, the components are relatively inexpensive. The eye-tracking bar cost is rapidly decreasing as the market for such sensors grows. We have, with the aim of reproducibility, detailed the open-source software used and the calculations set up for precise results. Moreover, the recording of all the sensors is synchronized by a software developed by our team, RGBD-Sync, which is open-source and accessible to all. We hope that these initiatives will encourage our peers to use the same types of hardware and software so that we can easily compare and reproduce our future experiments.

Physiological Manifestations of reasoning

Our state of the art study, on the one hand, and the analysis of the features recorded during our experience, on the other hand, have allowed us to highlight the physiological signs and ocular features that are significant for the study of reasoning and situational awareness. We have, through three modalities, identified features that give crucial information on reasoning. The interest in observing these features is supported by many theories that we have put forward throughout this manuscript. The use of several modalities brings a double advantage:

1. Enhances robust detection: it is common to see a modality occulted by subject movement or poor sensor recording conditions;
2. Reasoning manifestations are caused by complex systems and can be interpreted in many ways. Observing several manifestations, from different modalities, can be correlated to strengthen the hypothesis put forward about the results.

A Cognitive Model for Observing Awareness

We believe that Artificial Intelligence and Human-Computer Interaction can benefit from results of Cognitive Science. We have developed a cognitive model to observe the Situation Awareness and knowledge of chess players engaged in problem-solving. This model aims at unifying elements of Artificial Intelligence and Cognitive Science. The different contributions of this model are:

1. **Knowledge Representation using Frames:** the data structures proposed by the Frames theory is easily implementable as a computer program. Representations of concepts, relationships, and a hierarchy of frames are also easily realizable. Program languages based on production rules can benefit from the simplicity of Frames and their detailed description.
2. **Working Memory Model:** the human memory used for reasoning is limited in storage space. It is essential to take this characteristic into account to interpret the reasoning steps of the subject correctly. Chunking depends on the subject's expertise and ability to perceive relevant information to solve a problem. The interactions between memory components, their different reaction times, and interactions with the external world should be taken into account in the observation of situational awareness.
3. **Evidence for Awareness from fixation:** Observations of manifestations of reasoning and situational awareness must be interpreted with the knowledge that these manifestations are arising from complex systems. The observation of a fixation on a significant element should not be interpreted as a complete understanding of that element, but should instead be seen as evidence of probabilistic consideration. The formulas that we propose make it possible to take this evidence into account.

8.3 Potential Applications

Collaborative Intelligent Systems

Agents and assistants can benefit from a model of reasoning to interpret human activities. For example, intelligent assistants mounted in transport systems, such as planes, cars, trucks, or buses, could effectively analyze the external manifestations of drivers and provide them with appropriate services. There are already assistants who ensure the safety of drivers based on their eye attention. Multi-modal agents based on our instrument could benefit from more robust and comprehensive recordings. Finally, situational awareness models could benefit collaborative robots, for example, in industry, to provide services and better respond to the needs of workers by observing them more efficiently.

Training and Education

Education technologies offer a natural application domain for technologies for monitoring attention, awareness, and comprehension. Developing tools capable of observing a student, identifying his errors, and directing him towards the right solution is a research topic that can benefit from the contributions made by this thesis. Indeed, from the data collected during our experiments, it is possible to conceive an intelligent software that can detect the errors of a chess player and ask him to apply the patterns learned by observing the experts. To generalize these behaviors to the field of general education would be great. An example of an intelligent tool incorporating the sensors we used is shown in Figure 8.1. This simple sketch details an intelligent student pulpit that would be able to observe and adapt the proposed content based on the user's reactions.

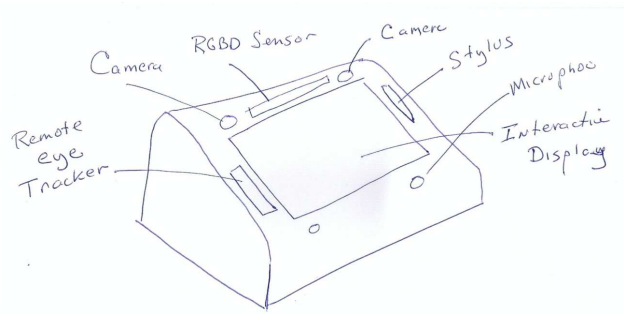


Figure 8.1: Student Aware intelligent Training Pulpit. Collaboration between the Pervasive team (LIG) and Philippe Dessus and colleagues at LaRAC

Also, simulators such as driving or pilot's simulators can be augmented with the sensors we offer to observe the physiological manifestations of the student in delicate situations.

Socially Aware Service Robots

These technologies can also be used to enhance interaction for social robotics. Humans communicate through a variety of non-verbal and paralinguistic channels. In order to achieve efficient human-centered systems, the model must be able to interpret these non-verbal cues. We have provided in this thesis many theories and features, covering different modalities, to capture and interpret these external signs. Pet robot or counter robot staff should be able to adapt to observable physiological reactions of the users to provide relevant services.

Ambient Intelligence

Smart homes are more and more equipped with several cameras that observe human activity to provide them relevant and appropriate services. A model of possible observable activities are, most of the time, fixed and rather small to ensure good performances. By combining existing sensors with the ones with are using in our project could increase the possible observable activities and improve the robustness of the system. Also, the mental model we proposed can add beneficial information to Smart Homes system to better understand human activities.

Bibliography

- Roe Admon, Gad Lubin, Orit Stern, Keren Rosenberg, Lee Sela, Haim Ben-Ami, and Talma Hendler. Human vulnerability to stress depends on amygdala's predisposition and hippocampal plasticity. *Proceedings of the National Academy of Sciences*, 106(33):14120–14125, 2009.
- Jonathan Aigrain, Michel Spodenkiewicz, Séverine Dubuiss, Marcin Detyniecki, David Cohen, and Mohamed Chetouani. Multimodal stress detection from multiple assessments. *IEEE Transactions on Affective Computing*, 9(4):491–506, 2016a.
- Jonathan Aigrain, Michel Spodenkiewicz, Severine Dubuisson, Marcin Detyniecki, David Cohen, and Mohamed Chetouani. Multimodal stress detection from multiple assessments. *IEEE Transactions on Affective Computing*, 2016b.
- Floyd Henry Allport. *Social psychology*. Boston, New York Houghton Mifflin, 1924.
- John R Anderson. A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, 22(3):261–295, 1983.
- Glenda Andrews and Graeme S Halford. A cognitive complexity metric applied to cognitive development. *Cognitive psychology*, 45(2):153–219, 2002.
- Michael Argyle. *Bodily communication*. 1988.
- Solomon E Asch. *Social psychology*. New York: Oxford University Press. (Original work published 1952), 1987.
- Richard C Atkinson and Richard M Shiffrin. Human memory: A proposed system and its control processes. In *Psychology of learning and motivation*, volume 2, pages 89–195. Elsevier, 1968.
- Alan Baddeley. Working memory. *Science*, 255(5044):556–559, 1992.
- Alan Baddeley. The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, 4(11):417–423, 2000.
- Alan D Baddeley and Graham Hitch. Working memory. In *Psychology of learning and motivation*, volume 8, pages 47–89. Elsevier, 1974.
- Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 59–66. IEEE, 2018.
- Raffaella Balzarini. *Approche cognitive pour l'intégration des outils de la géomatique en sciences de l'environnement: modélisation et évaluation*. PhD thesis, 2013.

- Lisa Feldman Barrett. Solving the emotion paradox: Categorization and the experience of emotion. *Personality and social psychology review*, 10(1):20–46, 2006.
- Lisa Feldman Barrett. Was darwin wrong about emotional expressions? *Current Directions in Psychological Science*, 20(6):400–406, 2011.
- Lisa Feldman Barrett, Batja Mesquita, and Maria Gendron. Context in emotion perception. *Current Directions in Psychological Science*, 20(5):286–290, 2011.
- Jon Barwise. *The situation in logic*. Number 17. Center for the Study of Language (CSLI), 1989.
- Antoine Bechara, Antonio R Damasio, Hanna Damasio, and Steven W Anderson. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50(1-3):7–15, 1994.
- Antoine Bechara, Hanna Damasio, and Antonio R Damasio. Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex*, 10(3):295–307, 2000.
- R. L. Birdwhistell. *Kinesics and context*. Philadelphia: University of Pennsylvania Press., 1970.
- DE Broadbent. Perception and communication. elmsford, ny, us, 1958.
- Donald E Broadbent, Peter FitzGerald, and Margaret HP Broadbent. Implicit and explicit knowledge in the control of complex systems. *British journal of Psychology*, 77(1):33–50, 1986.
- Roger Brown and James Kulik. Flashbulb memories. *Cognition*, 5(1):73–99, 1977.
- Tony W Buchanan. Retrieval of emotional memories. *Psychological bulletin*, 133(5):761, 2007.
- Walter B Cannon. The james-lange theory of emotions: A critical examination and an alternative theory. *The American journal of psychology*, 39(1/4):106–124, 1927.
- Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*, 2018.
- Peter R Chapman and Geoffrey Underwood. Visual search of driving situations: Danger and experience. *Perception*, 27(8):951–964, 1998.
- Neil Charness. Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance*, 7(2):467, 1981.
- Neil Charness, Eyal M Reingold, Marc Pomplun, and Dave M Stampe. The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & cognition*, 29(8): 1146–1152, 2001.
- William G Chase and K Anders Ericsson. Skilled memory. *Cognitive skills and their acquisition*, pages 141–189, 1981.
- William G Chase and Herbert A Simon. Perception in chess. *Cognitive psychology*, 4(1):55–81, 1973a.
- William G Chase and Herbert A Simon. The mind’s eye in chess. In *Visual information processing*, pages 215–281. Elsevier, 1973b.

- Micheline TH Chi. Quantifying qualitative analyses of verbal data: A practical guide. *The journal of the learning sciences*, 6(3):271–315, 1997.
- Nelson Cowan. *Attention and memory: An integrated framework*, volume 26. Oxford University Press, 1998.
- Nelson Cowan. An embedded-processes model of working memory. *Models of working memory: Mechanisms of active maintenance and executive control*, 20:506, 1999.
- Nelson Cowan. *Working Memory Capacity: Classic Edition*. Routledge, 2016.
- Kenneth James Williams Craik. *The nature of explanation*, volume 445. CUP Archive, 1952.
- Hewitt D Crane and Carroll M Steele. Generation-v dual-purkinje-image eyetracker. *Applied Optics*, 24(4):527–537, 1985.
- A Damasio. Somatic markers and the guidance of behavior/antonio r. damasio. NY: OUP, 1991.
- Antonio R Damasio. Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33(1-2):25–62, 1989.
- Antonio R Damasio. The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1346):1413–1420, 1996.
- Charles Darwin. The expression of the emotions in man and animals. London, UK: John Marry, 1872.
- A. D. De Groot. *Thought and choice in chess*. Walter de Gruyter GmbH & Co KG, 1978.
- RL De Valois and KK De Valois. Oxford psychology series, no. 14. *Spatial vision*. New York, NY, US: Oxford University Press, 1988.
- MJ Den Uyl and H Van Kuilenburg. The facereader: Online facial expression recognition. In *Proceedings of measuring behavior*, volume 30, pages 589–590, 2005.
- Thomas Dixon. *From passions to emotions: The creation of a secular psychological category*. Cambridge University Press, 2003.
- T D’orazio, M Leo, and A Distanto. Eye detection in face images for a driver vigilance system. In *IEEE Intelligent Vehicles Symposium, 2004*, pages 95–98. IEEE, 2004.
- Andrew T Duchowski. Eye tracking methodology. *Theory and practice*, 328(614):2–3, 2007.
- Richard O Duda, Peter E Hart, David G Stork, et al. *Pattern classification*, volume 2. Wiley New York, 1973.
- Claudia Ehmke and Stephanie Wilson. Identifying web usability problems from eye-tracking data. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 1*, pages 119–128. British Computer Society, 2007.
- Paul Ekman. A methodological discussion of nonverbal behavior. *The Journal of Psychology*, 43(1):141–149, 1957.

- Paul Ekman. Facial action coding system. 1977.
- Paul Ekman. *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. Henry Holt and Company, 2003.
- Paul Ekman and Wallace V Friesen. Nonverbal leakage and clues to deception. *Psychiatry*, 32(1): 88–106, 1969.
- Paul Ekman and Wallace V Friesen. Measuring facial movement. *Environmental psychology and nonverbal behavior*, 1(1):56–75, 1976.
- Paul Ekman and Dalai Lama. *Emotional Awareness: Overcoming the Obstacles to Psychological Balance and Compassion*. Henry Holt and Company, 2008.
- Paul Ekman and Wallace V. Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17:124–9, 03 1971. doi: 10.1037/h0030377.
- Ralf Engbert, André Longtin, and Reinhold Kliegl. A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision research*, 42(5):621–636, 2002.
- Randall W Engle, Stephen W Tuholski, James E Laughlin, and Andrew RA Conway. Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology: General*, 128(3):309, 1999.
- Gisela Erdmann and Beatrix Van Lindern. The effects of beta-adrenergic stimulation and beta-adrenergic blockade on emotional reactions. *Psychophysiology*, 17(4):332–338, 1980.
- K Anders Ericsson. Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts’ performance on representative tasks. *The Cambridge handbook of expertise and expert performance*, pages 223–241, 2006.
- K Anders Ericsson and Neil Charness. Expert performance: Its structure and acquisition. *American psychologist*, 49(8):725, 1994.
- K Anders Ericsson and Walter Kintsch. Long-term working memory. *Psychological review*, 102(2): 211, 1995.
- Norbert Freedman. The analysis of movement behavior during the clinical interview. *Studies in dyadic communication*, 7:153–175, 1972.
- Wallace V Friesen, Paul Ekman, et al. Emfacs-7: Emotional facial action coding system. *Unpublished manuscript, University of California at San Francisco*, 2(36):1, 1983.
- Nico H Frijda. *The emotions*. Cambridge University Press, 1986.
- Joaquin Fuster. *The Prefrontal Cortex Anatomy, Physiology and Neuropsychology of the Frontal Lobe*. 01 1997.
- Fernand Gobet and Herbert A. Simon. Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review*, 3(2):159–163, Jun 1996. ISSN 1531-5320. doi: 10.3758/BF03212414. URL <https://doi.org/10.3758/BF03212414>.
- Ellen Goeleven, Rudi De Raedt, Lemke Leyman, and Bruno Verschuere. The karolinska directed emotional faces: a validation study. *Cognition and emotion*, 22(6):1094–1118, 2008.

- E Bruce Goldstein. Sensation and perception. pacific grove, ca: Brooks. *Cole Publishing Company. Graesser, AC, & McMahan, C.(1993). Anomalous information triggers questions when adults solve quantitative problems and comprehend stories. Journal of Educational Psychology, 85(1): 136151, 1999.*
- Arthur C Graesser, Murray Singer, and Tom Trabasso. Constructing inferences during narrative text comprehension. *Psychological review, 101(3):371, 1994.*
- Karl Grammer, Bernhard Fink, Elisabeth Oberzaucher, Michaela Atzmüller, Ines Blantar, and Philipp Mitteroecker. The representation of self reported affect in body posture and body posture simulation. *Collegium antropologicum, 28(2):159–173, 2004.*
- James J Gross. Emotion regulation: Current status and future prospects. *Psychological Inquiry, 26(1):1–26, 2015.*
- Thomas Guntz, Raffaella Balzarini, Dominique Vaufreydaz, and James L Crowley. Multi-modal observation and interpretation of subjects engaged in problem solving. *arXiv preprint arXiv:1710.04486, 2017.*
- Thomas Guntz, Raffaella Balzarini, Dominique Vaufreydaz, and James Crowley. Multimodal observation and classification of people engaged in problem solving: Application to chess players. *Multimodal Technologies and Interaction, 2(2):11, 2018a.*
- Thomas Guntz, James L Crowley, Dominique Vaufreydaz, Raffaella Balzarini, and Philippe Dessus. The role of emotion in problem solving: first results from observing chess. In *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data*, page 12. ACM, 2018b.
- Graeme S Halford, William H Wilson, and Steven Phillips. Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21(6):803–831, 1998.*
- Robert G Harper, Arthur N Wiens, and Joseph D Matarazzo. *Nonverbal communication: The state of the art.* John Wiley & Sons, 1978.
- Jinni A Harrigan. Self-touching as an indicator of underlying affect and language processes. *Social science & medicine, 20(11):1161–1168, 1985.*
- Donald O Hebb. *The organization of behavior.* na, 1961.
- Donald Olding Hebb. *The organization of behavior*, volume 65. Wiley New York, 1949.
- Douglas L Hintzman. "schema abstraction" in a multiple-trace memory model. *Psychological review, 93(4):411, 1986.*
- Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. *Eye tracking: A comprehensive guide to methods and measures.* OUP Oxford, 2011.
- Panagiotis G Ipeirotis, Foster Provost, and Jing Wang. Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD workshop on human computation*, pages 64–67. ACM, 2010.

- David E Irwin. Visual memory within and across fixations. In *Eye movements and visual cognition*, pages 146–165. Springer, 1992.
- William James. II.—WHAT IS AN EMOTION ? *Mind*, os-IX(34):188–205, 1884. ISSN 0026-4423. doi: 10.1093/mind/os-IX.34.188. URL <https://doi.org/10.1093/mind/os-IX.34.188>.
- Wafa Johal, Damien Pellier, Carole Adam, Humbert Fiorino, and Sylvie Pesty. A cognitive and affective architecture for social human-robot interaction. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, pages 71–72. ACM, 2015.
- Wafa Benkaouar Johal. *Companion Robots Behaving with Style: Towards Plasticity in Social Human-Robot Interaction*. PhD thesis, 2015.
- P. N. Johnson-Laird. Foundations of cognitive science. chapter Mental Models, pages 469–499. MIT Press, Cambridge, MA, USA, 1989. ISBN 0-262-16112-5. URL <http://dl.acm.org/citation.cfm?id=102953.102965>.
- Daniel Kahneman. A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, 58(9):697, 2003.
- Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011.
- Madhu Kalia. Neurobiological basis of depression: an update. *Metabolism*, 54(5):24–27, 2005.
- Otto Klineberg. *Social psychology*. New York : Holt., 1940.
- Jaap M Koolhaas, Alessandro Bartolomucci, Bauke Buwalda, Seitse F de Boer, Gabriele Flügge, S Mechiel Korte, Peter Meerlo, Robert Murison, Berend Olivier, Paola Palanza, et al. Stress revisited: a critical evaluation of the stress concept. *Neuroscience & Biobehavioral Reviews*, 35(5):1291–1301, 2011.
- William R Kunst-Wilson and Robert B Zajonc. Affective discrimination of stimuli that cannot be recognized. *Science*, 207(4430):557–558, 1980.
- Hannu Kuusela and Pallab Paul. A comparison of concurrent and retrospective verbal protocol analysis. 113:387–404, 02 2000.
- W. Labarre. The cultural basis of emotions and gestures. *Journal of Personality*, 16:49–68, 1947.
- Oliver Langner, Ron Dotsch, Gijsbert Bijlstra, Daniel HJ Wigboldus, Skyler T Hawk, and AD Van Knippenberg. Presentation and validation of the radboud faces database. *Cognition and emotion*, 24(8):1377–1388, 2010.
- Richard S Lazarus. Psychological stress and the coping process. 1966.
- Richard S Lazarus. From psychological stress to the emotions: A history of changing outlooks. *Annual review of psychology*, 44(1):1–22, 1993.
- Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Trevino Robert, Jiliang Tang, and Huan Liu. Feature selection: A data perspective. *arXiv:1601.07996*, 2016.

- Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014. URL <http://arxiv.org/abs/1405.0312>.
- Julian Galindo Losada. *Adaptation des interfaces utilisateurs aux émotions*. PhD thesis, 2019.
- Hugo Lövheim. A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical hypotheses*, 78(2):341–348, 2012.
- Yuliya Lutchyn, Paul Johns, Mary Czerwinski, Shamsi Iqbal, Gloria Mark, and Akane Sano. Stress is in the eye of the beholder. pages 119–124, 09 2015. doi: 10.1109/ACII.2015.7344560.
- Paul D MacLean. *The triune brain in evolution: Role in paleocerebral functions*. Springer Science & Business Media, 1990.
- JL McGaugh. Amygdala: role in modulation of memory storage. *The amygdala: A functional analysis*, 2000.
- Gail McKoon and Roger Ratcliff. Inference during reading. *Psychological review*, 99(3):440, 1992.
- Albert Mehrabian and John T Friar. Encoding of attitude by a seated communicator via posture and position cues. *Journal of Consulting and Clinical Psychology*, 33(3):330, 1969.
- George A Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.
- George A Miller, Eugene Galanter, and Karl H Pribram. Plans and the structure of behavior. 1960.
- Marvin Minsky. A framework for representing knowledge. Technical report, Cambridge, MA, USA, 1974.
- Mortimer Mishkin, Leslie G Ungerleider, and Kathleen A Macko. Object vision and spatial vision: two cortical pathways. *Trends in neurosciences*, 6:414–417, 1983.
- Changiz Mohiyeddini and Stuart Semple. Displacement behaviour regulates the experience of stress in men. *Stress (Amsterdam, Netherlands)*, 16, 09 2012. doi: 10.3109/10253890.2012.707709.
- Agnes Moors. Theories of emotion causation: A review. *Cognition and emotion*, 23(4):625–662, 2009.
- Allen Newell, Herbert Alexander Simon, et al. *Human problem solving*, volume 104. Prentice-hall Englewood Cliffs, NJ, 1972.
- Charles E Osgood. Dimensionality of the semantic space for communication via facial expressions. *Scandinavian journal of Psychology*, 7(1):1–30, 1966.
- James W Papez. A proposed mechanism of emotion. *Archives of Neurology & Psychiatry*, 38(4): 725–743, 1937.
- Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8):1226–1238, 2005.

- John Perry and Jon Barwise. *Situations and attitudes*. MIT Press, 1983.
- Luiz Pessoa and Ralph Adolphs. Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. *Nature reviews neuroscience*, 11(11):773, 2010.
- Lloyd Peterson and Margaret Jean Peterson. Short-term retention of individual verbal items. *Journal of experimental psychology*, 58(3):193, 1959.
- Rosalind W Picard. *Affective computing*. MIT press, 2000.
- Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE transactions on biomedical engineering*, 58(1):7–11, 2010.
- Alex Poole and Linden J Ball. Eye tracking in hci and usability research. *Encyclopedia of human computer interaction*, 1:211–219, 2006.
- Maxime Portaz, Maxime Garcia, Adela Barbulescu, Antoine Begault, Laurence Boissieux, Marie-Paule Cani, Rémi Ronfard, and Dominique Vaufreydaz. Figurines, a multimodal framework for tangible storytelling author version. In *6th Workshop on Child Computer Interaction (WOCCI 2017) at the 19th ACM International Conference on Multi-modal Interaction (ICMI 2017)*., 2017.
- Jonathan Posner, James A Russell, and Bradley S Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(3):715–734, 2005.
- Gabriel A Radvansky and Rose T Zacks. The retrieval of situation-specific information. *Cognitive models of memory*, pages 173–213, 1997.
- Eyal M Reingold and Neil Charness. Perception in chess: Evidence from eye movements. *Cognitive processes in eye guidance*, pages 325–354, 2005.
- Eyal M Reingold, Neil Charness, Marc Pomplun, and Dave M Stampe. Visual span in expert chess players: Evidence from eye movements. *Psychological Science*, 12(1):48–55, 2001.
- Rainer Reisenzein. The schachter theory of emotion: Two decades later. *Psychological bulletin*, 94(2):239, 1983.
- James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.
- James A Russell. Is there universal recognition of emotion from facial expression? a review of the cross-cultural studies. *Psychological bulletin*, 115(1):102, 1994.
- James A Russell. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145, 2003.
- James A Russell and Albert Mehrabian. Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294, 1977.
- Stanley Schachter and Jerome Singer. Cognitive, social, and physiological determinants of emotional state. *Psychological review*, 69(5):379, 1962.

- K Scherer. Emotion. introduction to social psychology: A european perspective. m. hewstone and w. stroebe, 2000.
- Klaus R Scherer, Angela Schorr, and Tom Johnstone. *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press, 2001.
- Hans Selye. The physiology and pathology of exposure to stress. 1950.
- Hans Selye. Stress without distress. *Bruzelles médical*, 56:205–10, 06 1976. doi: 10.1007/978-1-4684-2238-2_9.
- Steven N Shapse. The diagnostic and statistical manual of mental disorders. 2008.
- Azim F Shariff and Jessica L Tracy. What are emotion expressions for? *Current Directions in Psychological Science*, 20(6):395–399, 2011.
- Heather Sheridan and Eyal M Reingold. Chess players’ eye movements reveal rapid recognition of complex visual patterns: Evidence from a chess-related visual search task. *Journal of vision*, 17(3):4–4, 2017.
- Dorothea P Simon and Herbert A Simon. Individual differences in solving physics problems. 1978.
- Herbert Simon and William Chase. Skill in chess. In *Computer chess compendium*, pages 175–188. Springer, 1988.
- Herbert A Simon and Michael Barenfeld. Information-processing analysis of perceptual processes in problem solving. *Psychological review*, 76(5):473, 1969.
- Herbert A Simon and Kevin Gilmarin. A simulation of memory for chess positions. *Cognitive psychology*, 5(1):29–46, 1973.
- Steven A Sloman. The empirical case for two systems of reasoning. *Psychological bulletin*, 119(1):3, 1996.
- Noldus Information Technology. FaceReader Methodolgy. 2015. URL http://www.noldus.com/files/file_manager/downloads/whitepaper/FaceReader_Methodology.pdf.
- OK Tikhomirov and ED Poznyanskaya. An investigation of visual search as a means of analyzing heuristics. *Soviet Psychology*, 5(2):3–15, 1966.
- S. S. Tomkins. *Affect, imagery, consciousness, Vol. 1. The positive affects*. New York, NY, US: Springer Publishing Co., 1962.
- Matthew J Traxler. *Introduction to psycholinguistics: Understanding language science*. John Wiley & Sons, 2011.
- Alfonso Troisi. Displacement activities as a behavioral measure of stress in nonhuman primates and human subjects. *Stress (Amsterdam, Netherlands)*, 5:47–54, 03 2002. doi: 10.1080/102538902900012378.
- Endel Tulving et al. Episodic and semantic memory. *Organization of memory*, 1:381–403, 1972.
- P Unema and M Rotting. Differences in eye movements and mental workload between experienced and inexperienced motor-vehicle drivers. david brogan: Visual search, 193-202, 1990.

- Dominique Vaufreydaz and Amaury Nègre. Mobilergbd, an open benchmark corpus for mobile rgb-d related algorithms. In *13th International Conference on Control, Automation, Robotics and Vision*, 2014.
- Dominique Vaufreydaz, Wafa Johal, and Claudine Combe. Starting engagement detection towards a companion robot using multimodal features. *Robotics and Autonomous Systems*, page 25, January 2015. doi: 10.1016/j.robot.2015.01.004. URL <https://hal.inria.fr/hal-01122396>.
- Peter C Wason and J St BT Evans. Dual processes in reasoning? *Cognition*, 3(2):141–154, 1974.
- Omri Weisman, Mohamed Chetouani, Catherine Saint-georges, Nadege Bourvis, Emilie Dela-herche, Orna Zagoory-Sharon, and David Cohen. Dynamics of non-verbal vocalizations and hormones during father-infant interaction. *IEEE Transactions on Affective Computing*, pages 1–1, 01 2015. doi: 10.1109/TAFFC.2015.2478468.
- Jacob Whitehill, Ting-fan Wu, Jacob Bergsma, Javier R Movellan, and Paul L Ruvolo. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in neural information processing systems*, pages 2035–2043, 2009.
- Jacqueline Wijsman, Bernard Grundlehner, Hao Liu, Hermie Hermens, and Julien Penders. Towards mental stress detection using wearable physiological sensors. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1798–1801. IEEE, 2011.
- Wikipedia contributors. Awareness — Wikipedia, the free encyclopedia. <https://en.wikipedia.org/w/index.php?title=Awareness&oldid=939302413>, 2020a. [Online; accessed 1-March-2020].
- Wikipedia contributors. Glossary of chess — Wikipedia, the free encyclopedia, 2020b. URL https://en.wikipedia.org/w/index.php?title=Glossary_of_chess&oldid=934436624. [Online; accessed 11-January-2020].
- Lowe Bryan William and Noble Harter. Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological review*, 6(4):345, 1899.
- Xiang Xiang and Trac Tran. Linear disentangled representation learning for facial actions. *IEEE Transactions on Circuits and Systems for Video Technology*, PP, 01 2017. doi: 10.1109/TCSVT.2017.2771150.
- Louis Émile Javal. *Essai sur la physiologie de la lecture*, 1879.

Appendix A

Retrieve depth from 2D keypoints and Kinect's recording

Once a 2D estimation (keypoints) has been extracted from the RGB video, the corresponding depth information can be retrieved from the 3D stream recorded by the Kinect. First, a linear transformation $\mathbb{R}^2 \rightarrow \mathbb{R}^3$ is performed on the RGB point. Let $p = \begin{bmatrix} x \\ y \end{bmatrix}$ the output 2D keypoint from OpenPose, $q = \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix}$ its equivalent in 3D and the 2×3 matrix M :

$$M = \begin{pmatrix} 2.931321973214652 & -0.009141788470755974 & 240.8456158271623 \\ -0.00231576534461303 & 2.905782928430264 & -55.02611560517637 \end{pmatrix}$$

Then, the linear transformation M can be applied on p :

$$p^T \cdot M = [x, y] \cdot M = [x', y', z'] \quad (\text{A.1})$$

The corresponding depth can retrieve in the look up table from the Kinect recording.

$$dz \leftarrow \text{depth.get}[x' + \text{Kinect.Width} \times y'] \times f \quad (\text{A.2})$$

f being a rescaling factor (set to $0.001f$ for our settings). Once the depth component (dz) has been retrieved, dx and dy are computed using the original point and constant parameters from Kinect settings:

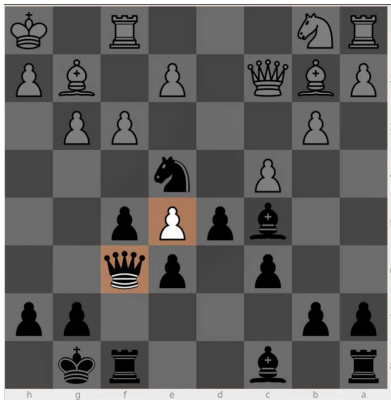
$$\begin{aligned} dx &= (x - ox) \times dz \times \text{inv_fx} \\ dy &= -(y - oy) \times dz \times \text{inv_fy} \end{aligned}$$

with:

$$\begin{aligned} \text{inv_fx} &= 1.f / 364.478f & ox &= 252.954f \\ \text{inv_fy} &= 1.f / 364.478f & oy &= 205.879f \end{aligned}$$

Appendix B

Relevant Chunks of a Task used in Experiment 2

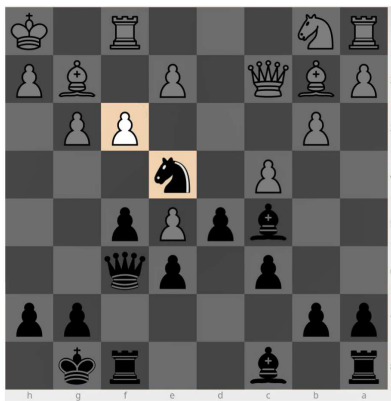


(a) Chunk 1: White Pawn E5 Threatens Black Queen F6

Chunk 1: White Pawn E5 Threatens Black Queen F6

```
(Chunk ChessChunk (C1)
  (Entities (WP5*), (BQ*))
  (Color (White))
  (Relation (Threatens))
  (Spatial-Extent (E5, F6)))
```

Entity **WP5** is the chess piece White Pawn E5.
Entity **BQ** is the chess piece Black Queen.

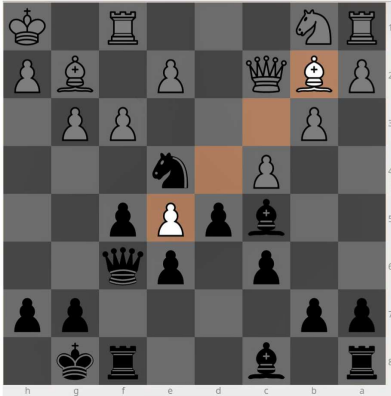


(b) Chunk 2: White Pawn F3 Threatens Black Knight E4

Chunk 2: White Pawn F3 Threatens Black Knight E4

```
(Chunk ChessChunk (C2)
  (Entities (WP3*), (BK1*))
  (Color (White))
  (Relation (Threatens))
  (Spatial-Extent (F3, E4)))
```

Entity **WP3** is chess piece White Pawn in F3.
Entity **BK1** is chess piece Black Knight in E4.

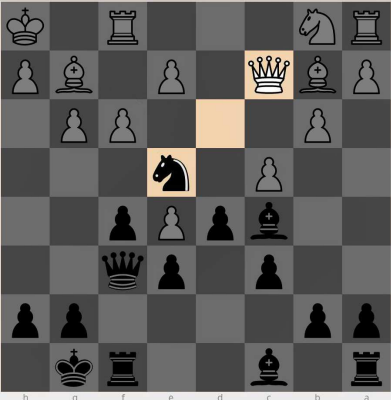


(a) Chunk 3: White Bishop B2 covers White Pawn E5

Chunk 3: White Bishop B2 covers White Pawn E5

```
(Chunk ChessChunk (C3)
  (Entities (WB2*), (WP5*))
  (Color (White))
  (Relation (Protects))
  (Spatial-Extent (B2, C3, D4, E5)))
```

Entity **WB2** is chess piece White Bishop in B2.
Entity **WP5** is chess piece White Pawn in E5.

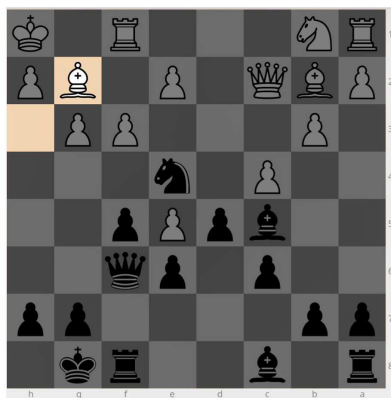


(b) Chunk 4: White Queen C2 Threatens Black Knight E4

Chunk 4: White Queen C2 Threatens Black Knight E4

```
(Chunk ChessChunk (C4)
  (Entities (WQ*), (BK1*))
  (Color (White))
  (Relation (Threatens))
  (Spatial-Extent (C2, D3, E4)))
```

Entity **WQ** is chess piece White Queen in C2.
Entity **BK1** is chess piece Black Knight in E4.

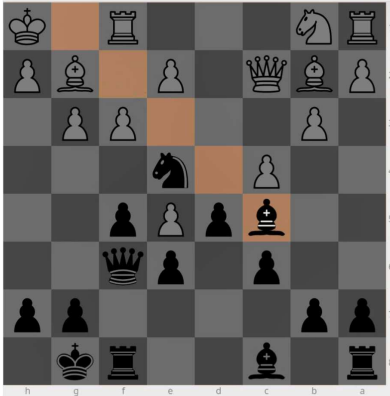


(c) Chunk 5: White Bishop G2 covers H3

Chunk 5: White Bishop G2 covers H3

```
(Chunk ChessChunk (C5)
  (Entities (WB1*), (SH3*))
  (Color (Black))
  (Relation (Controls))
  (Spatial-Extent (G2, H3)))
```

Entity **WB1** is chess piece White Bishop in G2. Entity **SH3** is the empty square H3.

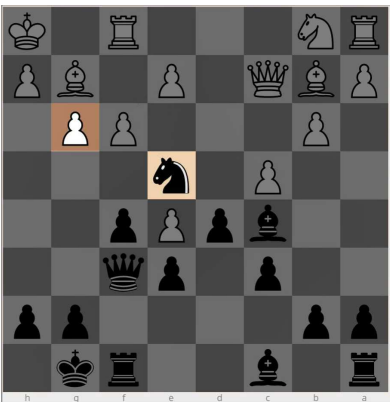


Chunk 6: Black Bishop C5 Controls Diagonal C5-G1

```
(Chunk ChessChunk (C6)
  (Entities (BB1*), (BDiag_1*))
  (Color (Black))
  (Relation (Controls))
  (Spatial-Extent (C5, D4, E3, F2, G1)))
```

Entity **BB1** is chess piece Black Bishop in C5.
Entity **BDiag_1** is the diagonal A7-G1.

(d) Chunk 6: Black Bishop C5 controls Diagonal C5-G1

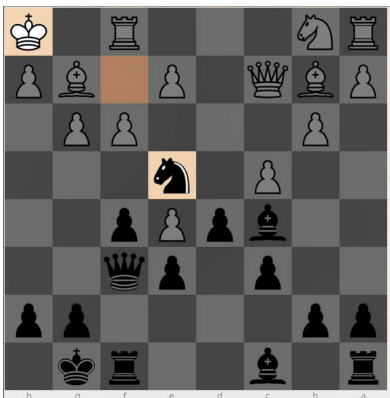


Chunk 7: Black Knight E4 threatens White Pawn G3

```
(Chunk ChessChunk (C7)
  (Entities (BK1*), (WP2*))
  (Color (Black))
  (Relation (Threatens))
  (Spatial-Extent (E4, F4, F3, G3)))
```

Entity **BK1** is chess piece Black Knight in E4.
Entity **WP2** is chess piece White Pawn in G3.

(e) Chunk 7: Black Knight E4 threatens White Pawn G3

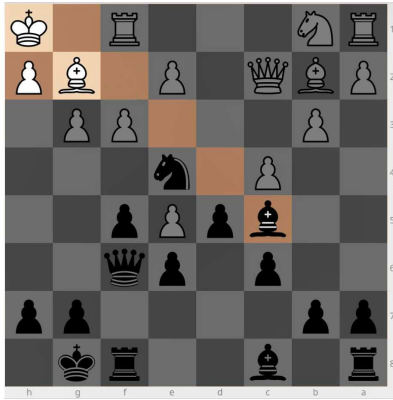


Chunk 8: Moving Black Knight E4 to F2: Check Attack, Sacrifice

```
(Chunk ChessChunk (C8)
  (Entities (BK1*), (SF2*), (WK*))
  (Color (Black))
  (Relation (Threatens))
  (Spatial-Extent (E4, F2, H1)))
```

Entity **BK1** is chess piece Black Knight in G2.
Entity **WK** is chess piece White King in H1.
Entity **SF2** is the empty square F2.

(f) Chunk 8: Moving Black Knight E4 to F2: Check Attack, Sacrifice



Chunk 9: White King H1 Trapped

```
(Chunk ChessChunk (C9)
  (Entities (WK*), (BDiag_1*), (SH2*), (SG2*))
  (Color (White))
  (Relation (Trapped))
  (Spatial-Extent (C5, D4, E3, F2, G1, G2, H2, H1)))
```

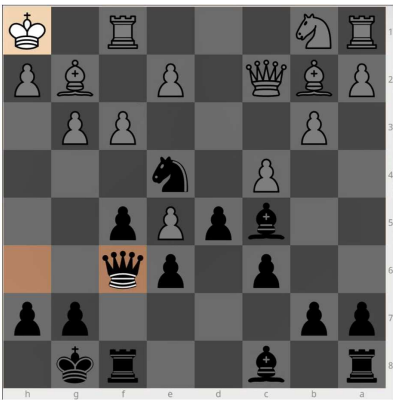
Entity **BB1** is chess piece Black Bishop in C5.

Entity **SH2** is the square H2.

Entity **SG2** is the square G2.

Entity **BDiag_1** is the diagonal A7-G1.

(g) Chunk 9: White King H1 Trapped



Chunk 10: Moving Black Queen F6 to H6: Check Attack

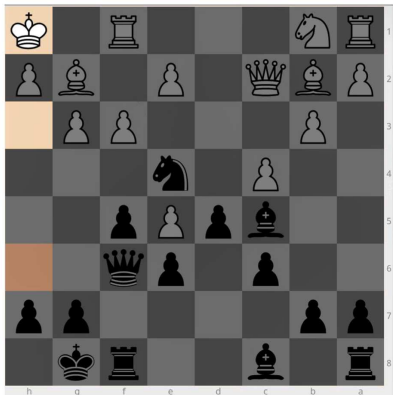
```
(Chunk ChessChunk (C10)
  (Entities (BQ*), (SH6*), (WK*))
  (Color (Black))
  (Relation (Threatens))
  (Spatial-Extent (F6, H6, H1)))
```

Entity **BQ** is chess piece Black Queen in F6.

Entity **SH6** is the empty square H6.

Entity **WK** is chess piece White King in H1.

(h) Chunk 10: Moving Black Queen F6 to H6: Check Attack



Chunk 11: Black Queen H6 moving to H3: Mate Attack, Sacrifice

```
(Chunk ChessChunk (C11)
  (Entities (BQ*), (SH3*), (WK*))
  (Color (Black))
  (Relation (Threatens))
  (Spatial-Extent (H6, H3, H1)))
```

Entity **BQ** is chess piece Black Queen in F6.

Entity **WK** is chess piece White King in H1.

Entity **SH3** is the empty square H3.

(i) Chunk 11: Black Queen H6 moving to H3: Mate Attack

Figure B.3: Relevant chunks of N-Check-Mate ($N = 3$) Task presented using Frames.

Appendix C

Publications

- Guntz, T., Balzarini, R., Vaufreydaz, D., & Crowley, J. L. (2017). Multimodal observation and interpretation of subjects engaged in problem solving. arXiv preprint arXiv:1710.04486.
- Guntz, T., Balzarini, R., Vaufreydaz, D., & Crowley, J. L. (2018). Multimodal observation and classification of people engaged in problem solving: Application to chess players. *Multimodal Technologies and Interaction*, 2(2), 11.
- Guntz, T., Crowley, J. L., Vaufreydaz, D., Balzarini, R., & Dessus, P. (2018, October). The role of emotion in problem solving: first results from observing chess. In *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data* (p. 12). ACM.
- Louedec, J. L., Guntz, T., Crowley, J. L., & Vaufreydaz, D. (2019). Deep learning investigation for chess player attention prediction using eye-tracking and game data. arXiv preprint arXiv:1904.08155.