



HAL
open science

De l'énactivisme appliqué à la mémoire humaine : Athena, un modèle fractal de covariances sensorimotrices

Johan Briglia

► To cite this version:

Johan Briglia. De l'énactivisme appliqué à la mémoire humaine : Athena, un modèle fractal de covariances sensorimotrices. Psychologie. Université Paul Valéry - Montpellier III, 2017. Français. NNT : 2017MON30076 . tel-01818765

HAL Id: tel-01818765

<https://theses.hal.science/tel-01818765>

Submitted on 19 Jun 2018

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

Délivrée par l' Université Paul-Valéry

Préparée au sein de l'école doctorale 60
Et de l'unité de recherche EPSYLON EA 4556
Équipe DynaCSE

Spécialité : **Psychologie Cognitive**

Présentée par : **Johan BRIGLIA**

**DE L'ÉNACTIVISME APPLIQUÉ
À LA MÉMOIRE HUMAINE**
ATHENA : un modèle fractal de covariances sensorimotrices

Soutenue le 18 Décembre 2017 devant le jury composé de

Pr. Denis BROUILLET, HDR, Université Paul-Valéry	Directeur de thèse
Dr. Lionel BRUNEL, HDR, Université Paul-Valéry	Examinateur
Pr. Fabrice GUILLAUME, HDR, Université d'Aix-Marseille	Rapporteur
Dr. Sophie MARTIN, HDR, Université Paul-Valéry	Examinatrice
Dr. Déborah NOURRIT, MCF, Université Paul-Valéry	Examinatrice
Pr. Véronique QUAGLINO, HDR, Université de Picardie	Présidente
Dr. Stéphane ROUSSET, MCF, Université Grenoble Alpes	Examinateur
Pr. Rémy VERSACE, HDR, Université Lyon 2	Rapporteur



Résumé

Depuis la révolution cogniviste des années 50, le paradigme cognitiviste a été accepté par la grande majorité des chercheurs, en particulier dans le domaine de la mémoire. Le connexionnisme est venu remettre en cause son organisation modulariste en mettant en avant son caractère émergent. Cependant, la question de la référence restait posée en ce sens qu'elle reposait sur l'existence de représentations internes d'un monde externe au sujet. Le paradigme énaïviste qui reprend les principes du constructivisme propose que la cognition ne soit plus considérée comme la représentation d'un monde préexistant mais comme le produit de nos expériences sensori-motrices. Dans ce travail, nous présentons ATHENA : un modèle de mémoire énaïviste sans représentations inspiré de MINERVA2 mais prenant en compte les apports du modèle Act-In. La mémoire se construit ici par une contextualisation fractale des covariances sensorimotrices. Une série de simulations expérimentales nous permet de rendre compte de nombreux phénomènes mnésiques de manière plus efficace que ce que permet le représentationnalisme : l'émergence, l'apprentissage et manipulation d'abstractions sans représentations ; mais aussi d'autres phénomènes cognitifs tels que la construction temporelle subjective (sans représentations temporelles apprises).

Mots-clés : mémoire, énaïvisme, représentation, modélisation, fractal

Since the cognitive revolution of the 1950s, the cognitive paradigm has been accepted by the great majority of researchers, especially in the domain of memory. Connectionism has come to question its modular organization by emphasizing its emergent character. However, the question of reference remained posed in the sense it was based on the existence of internal representations of a world external to the subject. The enactivist paradigm, that takes up the principles of constructivism, proposes that cognition should no longer be seen as the representation of a pre-existing world but as the product of our sensorimotor experiences. In this work, we present ATHENA : an enactivist model of memory without representations inspired by MINERVA2, but taking into account the contributions of the Act-In model. Memory is built by a fractal contextualization of sensorimotor covariances. A series of experimental simulations allows us to account for many memory phenomena in a more effective way than what is possible within representationalism : the emergence, learning and manipulation of abstractions without representations ; but also other cognitive phenomena such as a subjective time construction (without temporal representations learned).

Keywords : memory, enaction, representation, modeling, fractal

Remerciements

Je tiens à remercier tout particulièrement Denis Brouillet, sans qui je n'aurais pas eu l'opportunité d'en arriver là, et grâce à qui j'ai eu l'ambition de me construire en tant que chercheur. Je lui suis reconnaissant, entre autres, de ses nombreux enseignements, de l'espace de réflexion qu'il a ouvert au-delà du cadre traditionnel de la transmission unidirectionnelle, de la confiance contagieuse en mon travail dont il a témoigné, comme des critiques constructives qui m'ont permis de m'améliorer. Je lui suis infiniment redevable de nous avoir permis la co-construction d'un terrain ontologique sur lequel développer mes ambitions épistémologiques, ce fut une chance dont je prends toute la mesure. Les nombreux échanges et précieux conseils qui ont ponctué notre longue collaboration ont fait de lui une inspiration et un véritable exemple à suivre.

Je remercie sincèrement les membres de mon jury de thèse pour m'avoir fait l'honneur d'évaluer mon travail. Merci à Fabrice Guillaume et Rémy Versace d'avoir accepté d'être rapporteurs et merci à Lionel Brunel, Sophie Martin, Déborah Nourrit, Véronique Quaglino et Stéphane Rousset d'avoir accepté d'être examinateurs.

Je remercie également mes camarades doctorants et chercheurs qui m'ont accompagné dans cette épopée doctorale. Le développement théorique et mathématique de l'article principal autour duquel tourne ce travail n'aurait pas été possible sans l'aide considérable de Philippe Servajean, Arthur-Henri Michalland et Lionel Brunel. Merci également à Fabrice Da Silva pour son aide à la passation de simulations, et merci à Layla El Mossadeq pour le décryptage de l'infimum. Leur collaboration comme leur amitié me sont très chers.

Merci à Nicolas Petton pour m'avoir fait découvrir le langage Smalltalk et m'avoir appris à concrétiser informatiquement n'importe quel concept théorique. Merci également pour son aide combinée à celle de Camille Teruel, lesquelles m'ont permis d'effectuer un refactoring du logiciel ATHENA tellement efficace que sa vitesse d'exécution fut multipliée par 196.

Je tiens à remercier Stéphane Ducasse et toute l'équipe de recherche RMoD de l'INRIA-Lille pour m'avoir permis de présenter mes travaux et de mettre en place une collaboration interdisciplinaire nécessaire à la validation du modèle informatique d'ATHENA. Merci également à David Chisnall et à toute la communauté Smalltalk pour m'avoir permis de participer à l'organisation d'ESUG-Cambridge et d'amorcer une diffusion internationale de ce travail.

Je remercie sincèrement ma sœur et mes amis pour m'avoir toujours encouragé malgré mon capharnaüm théorique tout au long de cette aventure, et pour avoir fourni une relecture approfondie de ce travail. Merci à ceux qui furent à mes côtés, et merci à ceux qui le sont toujours.

Mais je dédie surtout cette thèse à mes parents. Pour m'avoir donné le choix qu'ils n'ont jamais eu, et pour m'avoir toujours soutenu et aidé de manière inconditionnelle. Nous sommes issus de l'interaction avec notre environnement, et ils ont su me fournir l'environnement et les interactions optimales nécessaires à ma construction personnelle comme à l'accomplissement de ce travail de longue haleine. Ce manuscrit est aussi le fruit de leurs apports inestimables.

« , , “ - ”.
, , , , .
, , .

À l'origine fut la vitesse, le pur mouvement furtif, le “vent-foudre”.
Puis le cosmos décéléra, prit consistance et forme, jusqu'aux lenteurs
habitables, jusqu'au vivant, jusqu'à vous.
Bienvenue à toi, lent Homme lié, poussif tresseur des vitesses.

ı` , .
Nous sommes , don .
de l'étoffe
faits de son tissé .
de , vents
ı` Nous sommes faits de l'étoffe dont sont tissés les vents. »

Alain DAMASIO
La Horde du Contrevent

Sommaire

Introduction	1
1 État de l'art	3
1.1 Considérations théoriques	3
1.1.1 A propos des représentations	3
1.1.1.1 Les représentations cognitivistes	3
1.1.1.2 Les représentations connexionnistes	7
1.1.2 Enactivisme sans représentation	10
1.2 Une mémoire sans représentation	13
1.2.1 Rejet de toute forme de computo-symbolisme	14
1.2.2 Le problème du contenu mnésique	16
1.2.3 Contextualisation et re-présentation	20
1.2.4 Meta-contextualisation de covariances fractales	24
1.2.5 Mémoire et construction temporelle	28
1.3 Conclusion et problématique	32
2 Modélisation	34
2.1 Modéliser la mémoire	34
2.2 Le modèle MINERVA2	36
2.2.1 Présentation de MINERVA2	36
2.2.2 Modifications nécessaires pour modéliser une mémoire sans repré- sentations	39
2.3 ATHENA : a fractal model of human memory	41
3 Simulations	63
3.1 ATHENA et MINERVA2	63
3.1.1 Rappel indicé	64
3.1.2 FURIG et Prototypie	67
3.1.3 FURIG et Réaction à la nouveauté	70
3.1.4 Memory-Conjunction Errors	74
3.1.5 Conclusion	77
3.2 Comportement temporel	78
3.2.1 Interférence relative	79
3.2.2 Effet de Primauté/Récence	82

3.2.3	Réverbération mnésique	85
3.2.4	Prédiction mnésique	88
3.2.5	Conclusion	90
4	Discussion	92
4.1	Énactivisme radical	92
4.2	Fondamentaux d'une mémoire énéactiviste	95
4.3	Mise à l'épreuve d'une mémoire énéactiviste	97
4.4	Futurs travaux	103
4.5	Conclusion	106
	Bibliographie	107
A	Code source du modèle ATHENA	I
A.1	Licence	I
A.2	Code-source	II

Introduction

La modélisation de la mémoire humaine tente de comprendre le comportement et l'organisation mnésique, mais il est une question implicite à laquelle chaque modèle essaie de répondre à sa façon : Qu'est-ce que contient la mémoire ? Le « hard problem of content » (Hutto & Myin, 2012) n'est toujours pas résolu, mais il est généralement admis que la mémoire contient des représentations du monde extérieur (Fodor, 1981). En effet, il serait contre-intuitif de penser autrement : si nous interprétons le monde extérieur, notre mémoire ne pourrait contenir que des interprétations de ce monde, donc des représentations de celui-ci. Mais cette affirmation n'est pas neutre. Bien entendu, nous ne remettons pas en question le fait que la mémoire se base sur une interprétation, mais le postulat du monde extérieur au système mnésique pose des limitations épistémologiques importantes. Pour résoudre ces problèmes, nous allons analyser en profondeur ce que signifie réellement le fait de modéliser une mémoire qui se base sur des représentations, afin de comprendre ce que pourrait apprendre la mémoire s'il ne s'agissait pas de représentations. Une mémoire sans représentation est fondamentalement une mémoire énonciviste (Varela, 1993, 1995), et si le fait de modéliser à partir de représentations est la méthode la plus pratique et intuitive d'expliquer la mémoire, la mémoire sans représentation est, quant à elle, difficilement explicable et modélisable. À notre connaissance, il n'existe à ce jour aucune modélisation énonciviste d'une mémoire sans représentation. À travers ce travail nous allons tenter de remédier à ce manque épistémologique en construisant le modèle ATHENA. Dans cette perspective, nous mettrons en exergue quatre points fondamentaux qu'une mémoire sans représentation se doit de respecter.

Le premier point permettra de décrire comment une mémoire sans représentation peut fonctionner à partir de l'énonction. L'écphorie synergétique offre un paradigme extrêmement utile pour organiser l'information énonctive et son utilisation. ATHENA se basera alors sur MINERVA2, lequel est une formalisation mathématique de l'écphorie synergétique. Le deuxième point cherchera à définir ce que peut apprendre une mémoire sans représentation. Nous verrons que l'information d'un processus de covariance sensorimotrice est suffisante pour pouvoir modéliser une mémoire énonciviste. Le troisième point déterminera comment utiliser ces covariances sensorimotrices. Nous verrons que la contextualisation des covariances par inférence est capable de construire de l'information pertinente à la situation présente en fonction de l'histoire des covariances apprises. Le quatrième point permettra de comprendre comment la mémoire est capable d'interagir avec son propre traitement de la situation. En nous inspirant des apports du modèle Act-In (Versace et al., 2014) et de l'inter-dépendance de ses traces et composantes mnésiques,

nous parviendront à la présentation du fonctionnement fractal d'une mémoire sans représentation. Une conséquence importante de ce dernier point (une construction temporelle émergente) nous permettra de nous assurer de la cohérence des interactions internes du modèle. Ces quatre points feront l'objet d'un premier chapitre.

Nous serons ainsi en mesure de formaliser mathématiquement les fondements d'une mémoire énaïviste en appliquant à MINERVA2 un fonctionnement de type Act-In. Cette formalisation mathématique automatisée de manière informatique permettra la passation de répliquions de protocoles expérimentaux. De cette manière, nous serons en mesure de mettre à l'épreuve le modèle ATHENA en comparant son comportement à celui de son équivalent représentationnaliste (MINERVA2). La formalisation mathématique d'ATHENA sera présentée dans le deuxième chapitre.

La question des représentations est fondamentalement une question propre à la capacité d'interprétation, et pour ne pas provoquer d'incohérence théorique, se passer de représentations ne devrait pas poser de problèmes d'abstraction. Pour s'en assurer, nous allons présenter une première série de simulations expérimentales vérifiant la capacité d'ATHENA à faire émerger une abstraction, à apprendre une abstraction émergente, ainsi qu'à manipuler et confronter différentes abstractions émergentes. Cette série de simulations sera présentée dans l'article inclu dans le deuxième chapitre.

Une deuxième série de simulations nous assurera que le modèle ATHENA reproduit efficacement les apports du modèle MINERVA2. Se passer de représentations ne doit pas se faire en perdant les caractéristiques essentielles apportées par l'utilisation de représentations. Nous serons ainsi en mesure de montrer que l'utilisation de représentations n'est pas une nécessité.

Une troisième série de simulations nous permettra d'étudier le comportement temporel du modèle ATHENA. Si l'implémentation d'une mémoire énaïviste provoque une interdépendance totale entre les traces, le modèle devrait acquérir un comportement temporel émergent. De cette manière nous pourrions confirmer que l'ensemble des points fondamentaux propres à la modélisation sans représentation est efficacement modélisé. Les deux dernières séries de simulations seront présentées dans le troisième chapitre.

Ce document se terminera par une discussion portant sur les apports de ce travail à la lumière d'une conception énaïviste de la cognition. Cette discussion sera présentée dans le quatrième chapitre.

État de l'art

1.1 Considérations théoriques

1.1.1 A propos des représentations

Remettre en question la notion de représentation peut sembler secondaire dans l'étude de la mémoire. Certains trouveront que la question n'a pas lieu d'être posée. Ce serait comme remettre en question la notion de table : sur quoi pourrait-on manger si la notion de table n'existait pas ? Cette position cartésienne est généralement celle prise par les cognitivistes qui ne conçoivent pas que la cognition puisse effectuer des traitements sur autre chose que des représentations du monde. D'autres trouveront que remettre en question la notion de représentation est triviale. Bien entendu le cerveau ne contient pas un objet table, le traitement est distribué et c'est l'émergence de cette activité cognitive qui représente la table. Cette position connexionniste considère que les représentations sont un problème de terminologie méthodologique et qu'on parle de représentations car il serait trop compliqué d'explicitier la totalité des informations concernées.

Cependant, le recours à la notion de représentation n'est pas uniquement un problème d'ordre méthodologique, c'est un positionnement épistémologique lourd de conséquences. Une étude (brève) de l'historique de l'évolution de la notion de représentation dans le cadre des sciences cognitives nous permettra de mieux en comprendre son sens, et de proposer les modifications ontologiques et méthodologiques nécessaires pour une modélisation de la mémoire — et a fortiori de la cognition — qui s'en émancipe.

1.1.1.1 Les représentations cognitivistes

Le courant philosophique représentationnaliste est bien antérieur au cognitivisme. Cependant, comme le précisait Fodor (1981) la seule avancée apportée par le cognitivisme sur le représentationnalisme des XVIII^e et XIX^e siècles consiste en l'utilisation de l'ordinateur comme modèle cognitif.

L'ordinateur — plus exactement un ordinateur (i.e., un calculateur) — est un système structuré à l'extrême. Il est un aboutissement scientifique et culturel de l'application pratique de la logique humaine, et son extrême efficacité a donné au cognitivisme un argument d'autorité pour justifier son positionnement ontologique. En effet, ce n'est

pas le cognitivisme qui a donné naissance aux ordinateurs mais les ordinateurs qui ont donné naissance au courant cognitiviste (Varela, Thompson & Rosch, 1991). En soi, le fait de fonder une théorie sur des résultats techniques n'est pas un problème, mais à la condition de ne pas considérer son efficacité pratique comme un gage épistémologique. Pour le comprendre il est nécessaire de mettre à l'épreuve les théories sous-jacentes au structuralisme computationnel.

Comparer la cognition à un ordinateur pose un problème de cohérence : l'ordinateur fonctionne sur le principe de traitement de symboles — symboles qui n'entretiennent aucun lien avec ce qu'ils représentent ou ce qui les traite. Être indépendants de ce qui les traite conduit à distinguer le hardware (la machine) du software (le programme), autrement dit à considérer que la biologie du cerveau (le hardware) et la cognition (le software) sont indépendants. À vrai dire cette distinction entre hardware et software est à l'origine du problème que posent les représentations, et de manière plus globale du « hard problem of content », sur lequel nous reviendrons.

C'est au dualisme cartésien (Descartes, 1641) que remonte la distinction entre le hardware (le corps) et le software (l'esprit) du cognitivisme et la nécessité d'introduire l'existence de représentations. En effet, si le corps et l'esprit sont fondamentalement dissociés, comment l'esprit pourrait-il manipuler des concepts qui lui sont extérieurs ? La relation au corps s'apparentant à la relation au monde externe, il est nécessaire de faire appel à un processus de transduction qui permettrait de traduire l'externe vers l'interne (un modulateur en termes informatiques), ainsi que de traduire l'interne vers l'externe (un démodulateur), ce que fait un ordinateur. Mais on oublie trop vite qu'un ordinateur a eu besoin d'un technicien (i.e., un système totalement externe) pour définir les relations cohérentes du modem ou du bios. Or, faire appel à un technicien externe pour expliquer la construction de la cognition serait un fardeau épistémologique si lourd de conséquences qu'une autre solution doit être envisagée.

Si l'esprit traite des informations traduites de l'extérieur (et traduisibles vers l'extérieur), celui-ci est forcément un contenant manipulant des informations (i.e., des symboles) qui représentent les éléments externes. Dès lors, la notion de représentation est une nécessité ontologique de la dualité cartésienne, et non pas une théorie hors du champ computo-symbolique. Pourtant, même s'il devient largement admis que la distinction entre corps et esprit n'a pas lieu d'être (Damasio, 1994), le recours à la notion de représentation est loin d'être remis en question.

Cela peut s'expliquer par la cohérence épistémique que représente le cognitivisme fondé sur le dualisme cartésien qui a largement modelé la pensée occidentale. En effet, le cognitivisme se base sur trois postulats (Varela, 1993) qu'il convient de discuter pour éviter de reproduire leurs limites ontologiques.

En premier lieu, le cognitivisme postule que le monde est objectivable, qu'il est doué

de propriétés particulières telles que la longueur, la couleur, le mouvement, etc. Cette position d'une signification objective n'est pas toujours assumée telle quelle par les cognitivistes actuels qui reconnaissent un certain relativisme, mais nous verrons qu'elle est pourtant implicite au travers de leurs affirmations épistémologiques (« la science représente le monde tel qu'il est », « l'identité individuelle est un fait », etc.).

Deuxièmement, le cognitivisme postule que la cognition recompose ces propriétés objectives de manière plus ou moins fidèle à la réalité grâce à l'emploi de représentations internes. Il s'agit de la seule manière de faire le lien entre le monde externe et le sujet qu'impose le dualisme cartésien. Dès lors, remettre en question la seule notion de dualité cartésienne sans se séparer de ses conséquences représentationnalistes serait une erreur épistémologique.

Enfin, le cognitivisme postule que la notion de « soi » est dissociée du monde, et que la subjectivité effectue des traitements de manière indépendante. Ce postulat est l'autre versant du premier point développé ci-dessus, il permet d'expliquer l'internalité des processus cognitifs.

À ces trois postulats vient s'ajouter une des conséquences majeures du structuralisme : les recherches sur la cognition ne considère pas la cognition à partir du cadre de la cognition (Sartre, 1943 ; Dennet, 1978 ; Varela, 1993). Le chercheur cognitiviste se place au dehors de la cognition en oubliant ce fait essentiel qu'il est lui-même un système cognitif au cœur du monde qu'il étudie.

Concevoir la cognition comme un système qui reste indépendant de lui-même est directement issu de la dualité hardware-software des ordinateurs. Les opérations logiques effectuées par les transistors ne font pas partie des informations traitées par les ordinateurs (l'information existe mais elle n'est pas utilisée). Il en est tout autrement pour un système vivant, les opérations effectuées par la cognition font partie intégrante de la cognition. En effet, il n'existe pas de distinction fonctionnelle entre les neurones et les traitements qu'ils effectuent. À vrai dire, la manière dont l'information est traitée par les neurones (fréquence d'activation, et distribution d'activité) est la seule source d'information utilisée par ces mêmes neurones. À l'opposé, dans un système informatique l'ordinateur ne comprend pas ce qu'il fait, et s'il n'a pas de capacités réflexives c'est en partie parce qu'il ne se concentre que sur la forme des symboles qu'il traite et non sur leur sémantique. La signification échappe au symbole et ne peut être utilisée dans un ordinateur que si un informaticien donne un sens au symbole via les opérations logiques qui régissent le traitement des symboles.

Le cognitivisme est alors dans une position délicate : le problème de transduction entre le monde et le sujet, dont nous avons parlé plus haut, se retrouve aussi au cœur du sujet pensant. Une fois que l'information prétendument externe au sujet est potentiellement traduite en termes de représentations, comment le fait de manipuler des symboles pourrait-

il avoir un lien avec la signification de ces symboles ? Ce problème de transduction est récursif, et pour le résoudre il va falloir considérer que le symbole est sa propre sémantique de manière récursive (de la même manière que le sujet est sa propre interaction avec l'environnement).

Qu'en est-il alors de la conscience ? La notion de conscience est justement un problème central de la cognition qu'il ne faut pas reléguer à plus tard. En considérant que les processus cognitifs sont des computations de symboles, on affirme que la conscience — portée par les processus cognitifs — ne peut pas avoir accès au contenu de ces traitements qui permettent pourtant la conscience. Le cognitivisme a dès lors largement tendance à adhérer à la théorie de l'inconscient, mais une nouvelle distinction entre cognition et expérience est alors affirmée. Quel lien pourrait-il y avoir entre l'expérience vécue consciente et les processus computo-symboliques qui la provoquent ? Au delà de la dualité sujet-monde et symbolique-sémantique, le cognitivisme pose un nouveau problème de transduction entre cognition et expérience consciente. Ces trois problèmes de transduction provoquent une disparité des processus cognitifs, et les cognitivistes auront beau s'acharner à expliquer ces processus de manière méthodiques, ils n'auront jamais une vision globale du tableau psychologique pourtant nécessaire à sa compréhension.

« Vous entrez dans le cerveau par l'œil, vous remontez le nerf optique, vous tournez tout autour du cortex en regardant derrière chaque neurone, et avant même de vous en rendre compte, vous vous retrouvez dans la lumière du jour sur la crête d'une impulsion nerveuse motrice, vous grattant la tête et vous demandant où se trouve le soi. » (Dennett, 1984, p74-75)

En éloignant la cognition de la conscience, le cognitivisme affirme que la cognition est possible sans conscience. Or la conscience est la seule partie garante de l'ipséité, c'est-à-dire le sentiment que l'on a de rester la même personne au cours du temps qui passe malgré le fait qu'il s'opère constamment en nous des changements physiques et psychologiques. La conscience est en effet la seule forme possible du « soi ». En ayant adopté le modèle de l'ordinateur le cognitivisme affirme, en creux, que la cognition peut exister sans un « soi ». C'est pour cela que l'intelligence artificielle se concentre uniquement sur la question de la complexité des traitements pour avancer dans ses recherches, sans même considérer qu'il est possible de remettre en question la notion de « soi » du robot. Rien n'est jamais fait pour que le robot se considère comme une entité car c'est déjà censé être un donné, un absolu sur lequel l'IA se base. Ne pas résoudre les limites du cognitivisme est un frein à l'avancée de la recherche en sciences cognitives.

Pour pouvoir résoudre le problème de la conscience il faudrait déjà résoudre le problème de l'expérience vécue. Selon Jackendoff (1987) notre interprétation du monde se fait en fonction de notre expérience de celui-ci et non pas en fonction de son hypothétique nature. Nous percevons celui-ci comme tridimensionnel, macroscopique, coloré, et non pas dans un continuum d'espace-temps, en fonction d'ondes ou de particules élémentaires.

Notre cognition est fondamentalement construite en fonction d'un monde expérientiel (phénoménologiquement vécu). Mais dans ce cas, comment une cognition séparée de la conscience et du « soi » pourrait-elle porter sur notre expérience consciente du monde ? Pour répondre à cette question la compréhension de la cognition doit se baser sur la manière dont émerge la connaissance et non comment est traitée l'information.

1.1.1.2 Les représentations connexionnistes

La notion de représentation issue du dualisme cartésien, adoptée par le cognitivisme, est porteuse de deux limites techniques majeures qui trouvent leur origine dans la notion de traitement de l'information : la sérialité des processus et la localisation des symboles auxquelles le connexionnisme essayera de répondre mais, comme nous le verrons, de façon incomplète.

En premier lieu, l'utilisation de représentations demande que la computation soit linéaire (i.e., input, computation, output). Or, d'un point de vue biologique la latence des dendrites neurologiques ne permet qu'une sérialité limitée lors d'un temps de réponse rapide (une dizaine de neurones linéaires pour une réponse de 100ms).

En second lieu, la localisation des symboles d'un système représentationnel pose un problème de flexibilité psychologique incompatible avec la vicariance du cerveau. En effet, un système symbolique localisé rend le système trop fragile aux perturbations et autres modifications psychologiques constantes (liées au contexte ou à la dynamique interne de la cognition).

Contrairement au computo-symbolisme qui a adopté la métaphore de l'ordinateur pour expliquer la cognition, le connexionnisme a adopté la métaphore du cerveau, plus spécifiquement celle du fonctionnement cérébral. D'une part les neurones se comportent comme un système dynamique, mais on observe aussi une règle de réciprocité entre les activations apprises. Par exemple, si une région (un noyau, une couche) A se connecte à B, alors B se connecte réciproquement en retour à A (il n'y a que de rares exceptions mineures à cette règle). Il est par ailleurs important de remarquer, pour la suite, que ce type de coopérativité se retrouve à la fois au niveau local et au niveau global, laissant entrevoir une invariance d'échelle.

Par ailleurs, les logiques cognitivistes et connexionnistes diffèrent par leurs instruments mathématiques : si le cognitivisme utilise une mathématique discrète et un traitement séquentiel des processus cognitifs, le connexionnisme quant à lui utilise une mathématique continue pour calculer les variations qui se produisent en parallèle sur l'ensemble du réseau. De plus, il adopte une approche dynamique de la cognition, une logique qui s'éloigne de la séquentialité des concepts temporels régis par des règles absolues. Son évolution est gouvernée par un système d'équations différentielles qui formalise les variations interactives des différentes variables et l'instauration d'une configuration d'activation qui traduit la convergence du système d'équations vers une solution stable répondant aux conditions

initiales.

Enfin, au concept de représentations, le connexionnisme substitue le concept d'émergence. L'émergence est intimement liée à l'aspect dynamique d'un système complexe. En effet, dans un système complexe ce ne sont pas juste les entités distinctes du système qui sont porteuses d'information, mais le lien entre ces entités également. Si une entité A fait partie du même système complexe que l'entité B, l'information du système ne se limite pas à une superposition des entités A et B, mais comporte aussi l'interaction entre A et B. Dès lors le produit de l'interaction est plus que la somme de A et B, le tout est plus que la somme des parties (Smuts, 1926). Ce holisme ontologique ne se limite pas à contrecarrer le réductionnisme cognitiviste, mais il est surtout porteur d'une vision prometteuse de l'information cognitive : l'activité est une information. En conséquence de quoi, un même système complexe peut avoir des comportements différents en fonction des conditions et des interactions entre ses parties (Haken, 1983).

Qui plus est, la notion d'émergence donne toute son importance au contexte. Par exemple, l'état global des mêmes molécules d'eau peut se modifier drastiquement selon la température de ces molécules : les phénomènes solide, liquide ou gaz (ou autres) sont macroscopiquement différents malgré leur similarité structurelles de contenu. De la même manière, un système complexe peut avoir des comportements drastiquement différents avec le même contenu car contraint par le contexte.

Le contexte est ce qui conduit un ensemble de processus micro à avoir une logique macro, sans que les règles apparentes du système macro prennent racine dans de véritables règles micro (Smolensky, 1992). Par conséquent, le système vivant peut manifester des états fonctionnels émergents équivalents sans que la présence de représentations soit nécessaire : il suffit que son comportement soit contextuellement adapté à ces équivalences. Mais cela conduit à ne plus considérer le système cognitif comme distinct de son environnement.

Un déterminant majeur de l'émergence est le rôle que joue un agent sur lui-même. Un joueur d'échec expert n'a pas intériorisé toutes les possibilités de jeu et ne choisit pas le coup qui est objectivement le meilleur, il a plutôt intériorisé des règles supplémentaires qu'il s'est construit selon les situations rencontrées. Il est acteur, conjointement avec le jeu, de l'émergence car il est son propre sujet (le sujet du jeu est l'historicité des relations qu'il a eu avec ce jeu). L'émergence est ici une relation dynamique entre les différents traitements de jeu appris, et l'émergence de l'expertise ne se situe pas objectivement dans le joueur, mais dans la relation qu'il entretient avec le jeu. L'émergence n'est pas une computation, elle est une dynamique si elle veut être adaptable au contexte.

L'état d'un système connexionniste dépend des conditions initiales, de l'histoire de son fonctionnement et des situations rencontrées, tandis qu'un système computo-symbolique obéit à des règles rigides, indifférentes à la diversité des cas rencontrés et en nombre li-

mité. C'est en ce sens que le connexionnisme offre une alternative plus pertinente que le représentationnalisme computo-symbolique. Cependant, les émergences des systèmes connexionnistes se basent sur des parties élémentaires, qui sont, certes, plus pragmatiques que la représentation modulaire, mais qui restent des représentations. L'émergence du traitement de la couleur rouge se base sur des activations spécifiques à une longueur d'onde dans le sens où ces particularités représentent les particularités externes de la couleur rouge. Si le traitement connexionniste peut se passer de symboles, il se réfère plutôt à un niveau sub-symbolique pourtant porteur des mêmes limites que les représentations. Qu'est-ce qui permet de porter les premières formes sub-symboliques ? Qu'est-ce qui remplit la fonction du concepteur ?

Si les tenants du connexionnisme avec les notions d'émergence globale et de traitements distribués ont été les premiers à critiquer la vision computo-symbolique de la cognition, ils ne rejettent pas pour autant la notion de représentation. « Pour les connexionnistes, une représentation consiste en la correspondance entre un tel état global émergent et des propriétés du monde ; elle n'est pas fonction de symboles particuliers » (Varela, 1993). Mais si la notion de symbole n'est plus nécessaire dans un système connexionniste, la notion de représentation n'est pourtant pas remise en question. Car, comme pour le computo-symbolisme, la cognition est indépendante du monde qu'elle « observe » comme une entité propre. Comme le computo-symbolisme, le connexionnisme n'a pas tranché le nœud gordien de la référence : le monde en soi vs le monde expérimenté.

En pratique le connexionnisme reste rattaché à une vision séquentialiste et computationaliste du symbolisme (pourtant rejetés par l'approche dynamique). Substituer les symboles cognitivistes par un pattern d'activation n'est pas suffisant, et les travaux connexionnistes gardent toujours une méthode de séquentialité dans le traitement effectué (i.e., input/traitement/output). La réciprocité d'un traitement entre zones A et B comme nous avons pu le voir n'est en pratique pas respectée (même pour les plus gros projets en cours : le *Human Brain Project* est une version étendue du *Blue Brain Project*, lequel se base sur la modélisation d'une colonne neurale, archétype du traitement séquentiel en couches de neurones).

Cette conception séquentialiste du connexionnisme ne se limite pourtant pas simplement à une tradition cognitiviste qui facilite la compréhension des systèmes complexes. Nous avons affaire à un choix épistémologique sur la nature temporelle du traitement de l'information (Port, Cummins & McAuley, 1995). En effet, considérer la dimension temporelle comme une base absolue sur laquelle se construit la cognition, sans prendre en compte toute sa dimension expérientielle, rencontre le même genre de limite apportée par le cognitivisme sur la notion d'un « soi » comme base absolue de la cognition. Le connexionnisme est parti dans la bonne direction en rejetant le computo-symbolisme du cognitivisme, mais il aurait été nécessaire de sortir complètement du représentationnalisme

dans sa dimension temporelle. Considérer que le traitement neuronal représente directement la dimension temporelle est une erreur ontologique : nous le verrons par la suite, la phénoménologie nous permettra de concevoir la temporalité comme une construction de l'activité cognitive en interaction avec elle-même.

Que ce soit pour le problème sub-symbolique du traitement, pour la correspondance représentationnelle des patterns d'activation, ou pour le problème temporel de la séquentialité pseudo-dynamique, c'est dans ses limites représentationnalistes que le connexionnisme échoue à proposer une alternative au cognitivisme. Parler de dynamique n'est pas suffisant pour se soustraire aux limites computo-symboliques du structuralisme. Le paradigme représentationnaliste ne permet pas de rendre compte du caractère autonome des êtres vivants et de leur système cognitif, mais surtout il ne prend pas en compte le fait que la cognition n'est pas distincte du monde qu'elle interprète. « Nous ne pouvons nous exclure du monde pour comparer son contenu avec ses représentations : nous sommes toujours immergés dans ce monde » (Varela, 1996, p.98).

1.1.2 Enactivisme sans représentation

Le contextualisme énaïvistique proposé par Varela (1988, 1993, 1995, 1996, 1997) est — à notre sens — la seule alternative pertinente proposée pour dépasser le représentationnalisme. Selon Chemero (2013), la distinction entre cognitivisme et contextualisme remonte à l'opposition entre le structuralisme de Wundt (1874) et le fonctionnalisme de James (1890). Plus qu'une différence de contenu théorique entre ces approches de la psychologie, c'est principalement la manière de concevoir ce que doit être la cognition qui alimente ce débat depuis plus d'un siècle. Pour simplifier, du côté des structuralistes les chercheurs se penchent sur la manière dont la cognition permet d'interagir avec l'environnement, tandis que les fonctionnalistes cherchent à comprendre comment l'interaction avec l'environnement permet la cognition. Cette distinction se retrouve tout au long de l'histoire de la psychologie, comme nous pouvons le voir en comparant les travaux d'auteurs comme Chomsky (1969) ou Fodor (1987) à ceux de Piaget (1936) ou Gibson (1982).

Si le fonctionnalisme de James a été mis sous l'éteignoir par la révolution structuraliste du computo-symbolisme, il intéresse un nombre de plus en plus grand de chercheurs dans le champ des sciences cognitives — soit dans son actualisation vers la psychologie écologique pour les chercheurs plus influencés par Gibson (Turvey, Shaw, Reed, & Mace, 1981), soit vers l'énaïction pour ceux étant plus influencés par la phénoménologie (Varela, Thompson, & Rosch, 1991 ; Thompson, 2007). C'est à l'énaïction que nous allons nous intéresser.

L'énaïction est un terme inventé par Varela¹ reposant sur un historique fortement lié

1. Bien que le terme « énaïctif » ait déjà été utilisé par Bruner (1966) pour décrire les connaissances acquises et manifestées par l'action

au constructivisme (e.g., Piaget, 1936 ; Poincaré, 1902 ; Goldstein, 1934) et à la phénoménologie (Heidegger, 1927 ; Merleau-Ponty, 1945 ; Sartre, 1943).

La problématique de base de l'énaction est de remettre au centre de la cognition l'expérience vécue dont le *computo-symbolisme* et le *connexionnisme* ne pouvaient pas rendre compte (Varela, 1996, 1997). En d'autres mots, l'énaction se pose comme une étude phénoménologique de l'expérience cognitive.

Adopter le point de vue de l'énaction revient à concevoir la cognition comme interdépendante de son environnement : elle est incarnée et située. Son ancrage constructiviste l'a conduite à rendre compte de l'émergence des structures cognitives à partir de processus incarnés sur l'histoire récurrente des interactions entre un individu et son environnement : c'est ce que Varela a appelé l'autopoïèse (Varela, Maturana & Uribe, 1974 Varela, 1979).

Les systèmes autopoïétiques se définissent par le simple fait d'être capable de perdurer dans le temps malgré toutes les modifications provoquées par les perturbations subies, qu'il s'agisse de celles du système ou de celles de son environnement. Dès lors, les systèmes autopoïétiques n'ont pas de but ni de fonction. Les notions de but et de fonction, comme celles de « input » et « output », appartiennent au domaine de description du système autopoïétique par un observateur qui considère le système au sein d'un milieu dont il fournit les déterminations (Rudrauf et al., 2003). L'input provoque des modifications du système auxquelles il doit s'adapter, et l'output n'est pas tant le but de l'input qu'une résonance de l'activité cognitive (rétablissant un état d'équilibre suite à la perturbation provoquée par l'activité cognitive) : le système n'a pas de finalité fonctionnelle, il ne cherche qu'à se stabiliser. De plus, l'output modifie aussi le système qui l'a produit, démontrant par là même que l'output est plus que la simple sortie du système, c'est aussi une entrée.

Comme nous l'avons dit plus haut, le positionnement épistémologique de l'énaction est fondamentalement non représentationnaliste. Pour l'expliquer, Varela (1993) a tenté de répondre à la question : « Qui est venu en premier, le monde ou l'image ? ». Il existe traditionnellement deux manières de répondre à cette question.

La première est la position objectiviste adoptée par le *computo-symbolisme* et le *connexionnisme* qui considèrent que le monde préexiste à l'image selon des règles qui lui sont externes. Il s'agit de « la position de la poule », elle est la plus intuitive mais est pourtant erronée puisque par exemple les catégories de couleurs relèvent de notre expérience vécue de ces couleurs (Hansen et al., 2006).

La seconde est la position subjectiviste qui considère que l'image provoque le monde selon des règles internes. Il s'agit de « la position de l'œuf », qui est tout autant erronée car les catégories de couleurs relèvent de nos particularités biologiques (e.g., le daltonisme) et sociologiques (e.g., la catégorie réunissant le jaune et le vert n'existe que dans une tribu amérindienne, Davidoff, 1999).

Cependant l'une comme l'autre relève d'une position représentationnaliste. En effet, la

première décrit une transduction de l'externe vers des représentations internes, tandis que la seconde décrit des projections de l'interne vers l'externe. Varela propose une troisième solution : la voie du milieu où le sujet et le monde se co-déterminent mutuellement. C'est cette voie dont rend compte l'énaction, laquelle ne fait pas de distinction entre le sujet et l'environnement (i.e., le sujet fait partie de l'environnement et l'environnement co-agit en parallèle du sujet). Nul besoin de représentations si le lien entre le sujet et l'environnement est continu.

Pour résumer l'énaction est aux antipodes de l'idée même de représentation : la conception représentationnaliste conçoit la cognition sous un schéma en trois temps (perception, traitement, action) où la perception et l'action sont distinctes du processus qui les traite ; tandis que pour la conception énative, perception et action sont interdépendants (ce en quoi consiste l'un dépend de ce en quoi consiste l'autre).

Si la perception peut orienter l'action, la perception est aussi la perception de la situation sensori-motrice liée à l'action qui en retour module la perception (le sensorimoteur est une rétro-activité dynamique perpétuelle). Dès lors, nous ne pouvons plus concevoir ce qui est perçu en termes d'événement ni même d'objet indépendant du processus de perception, mais en fonction de la nature sensorimotrice du système perceptif : nous percevons la manière dont nous traitons l'information. C'est, selon Brooks (1991), le principal obstacle que rencontre la recherche en intelligence artificielle : la représentation n'est pas la bonne unité d'abstraction dans la construction des parties les plus élaborées des systèmes intelligents. Pour modéliser un système cognitif, il faut lui faire apprendre la manière dont il traite l'information.

Nous ne sommes pas réductibles à un cerveau isolé à côté d'un monde propre, nous sommes des entités dynamiques en lien constant avec notre environnement (Cosmelli & Thompson, 2010). L'environnement fait partie du système cognitif. Dès lors, la cognition ne traite pas des représentations du monde extérieur, mais de ce lien direct qui est déjà en soi de l'information, une information bien concrète (Varela, 1995).

Dès lors, nous serions tentés de rejoindre Menary (2010) qui considère qu'une rupture totale doit être effectuée entre le cognitivisme et l'énactivisme car la relation arbitraire qu'établit le cognitivisme entre physique, sémantique et symbole doit être rejetée dans sa totalité.

En effet, l'énactivisme ne doit pas être une manière de moderniser le cognitivisme. La cognition n'est pas à proprement parler ni incarnée ni située, car elle fait partie intégrante du système sensori-moteur et de l'environnement. En d'autres mots, il n'y a pas d'esprit, juste une interaction dynamique entre un corps et un environnement et le cerveau ne traite que ces interactions à un niveau basique mais dynamiquement complexe.

Construire une compréhension de la cognition basée uniquement sur des covariances sensorielles et motrices contingentes pourrait sembler au premier abord une régression

scientifique puisqu'elle demande de ne plus se fier aux apports (parfois très efficaces) du cognitivisme. En effet, les covariances sensorielles et motrices contingentes ressemblent à une version améliorée de l'apprentissage associatif pour lequel Hebb (1949) a montré qu'au niveau cérébral il correspondait à la modification des connexions synaptiques entre neurones liée à l'activité même des neurones. Par conséquent, le cerveau a la capacité de modifier par lui-même la connectivité de ses neurones : c'est un système dynamique auto-organisé. Or, le caractère dynamique et auto-organisé du cerveau impliquant que les processus modifient en permanence le fonctionnement même du cerveau, nos perceptions se modifient elles aussi en permanence et par voie de conséquence le monde. Plus précisément, dans la mesure où le monde se transforme constamment à la suite de l'activité neuronale, la référence ne peut pas être dans un monde pré-donné, elle est dans la structure sensori-motrice du sujet en interaction avec l'environnement, ce que défend l'énaction et qui en fait sa modernité.

Dès lors pourquoi l'énaction n'est pas devenue le paradigme dominant des sciences cognitives en remplacement du cognitivisme ? Selon Di Paolo (2010), si la révolution énaïviste n'a pas encore eu lieu, c'est notamment lié à la difficulté de rendre compte de processus de haut niveau. Dès que les traitements cognitifs sont trop complexes pour être expliqués en simple termes sensorimoteurs, ils ont tendance à être expliqués en termes représentationnalistes (e.g., Agre & Chapman, 1987 ; Millikan, 1995 ; Clark & Toribio, 1994 ; Clark, 1997 ; Clark & Grush, 1999 ; Grush 2004 ; Chemero, 2013).

Le défi est donc de pouvoir modéliser un système énaïv sans représentation capable de rendre compte de situations expérimentales faisant appel à la fois à des traitements de bas et haut niveau. C'est ce que modestement va s'employer à montrer cette thèse à travers une modélisation fractale de la mémoire.

1.2 Une mémoire sans représentation

Dans un système structuraliste la mémoire est un module qui effectue une fonction précise dans le fonctionnement normal de la cognition, comme la perception par exemple. Dans un système énaïv la mémoire n'est pas distincte de la cognition : se souvenir de la manière dont on traite une information va permettre de réutiliser ce traitement. En d'autres mots, nous soutenons la thèse qu'une cognition énaïve est un système mnésique.

Nous avons pu mettre en exergue cinq critiques rédhibitoires de l'approche représentationnelle de la cognition (le *computo-symbolisme*, la dualité sujet-environnement, le *sub-symbolisme* représentationnel, la transduction récursive et la temporalité connexionniste) puis nous avons présenté l'énaction comme une réponse pertinente à ces critiques. Pour pouvoir s'assurer qu'une approche énaïve soit capable de résoudre ces limites, il est nécessaire de mettre ces théories à l'épreuve. Il n'existe, à notre connaissance, aucun modèle de mémoire sans représentation capable de rendre compte des mêmes capacités

que les modèles représentationnalistes. Nous allons dans ce chapitre décrire comment une approche énonciviste radicale de la cognition va nous permettre de mettre au point un modèle de mémoire sans représentation.

Dans un modèle sans représentation, le contenu des traces mnésiques ne peut être étudié indépendamment du reste du modèle. Ce qui peut sembler être une limite à sa compréhension n'est en fait pas un problème si on étudie ce modèle de la même manière qu'on étudie un sujet humain. À l'aide de simulations de protocoles expérimentaux déjà validés nous allons tester à quel point le modèle est conforme aux résultats obtenus chez des humains. Dès lors, ce n'est pas une compréhension objectivable des données contenues dans les traces qui est scientifiquement le sujet du modèle, mais bien les théories qui sont derrière : obtenir des résultats similaires aux résultats des sujets humains montrera que les théories énoncivistes sont capables de rendre compte de phénomènes mnésiques complexes sans utiliser la moindre forme de représentation.

1.2.1 Rejet de toute forme de *computo-symbolisme*

La critique la plus répandue du cognitivisme concerne les termes mêmes qui ont donné lieu à l'appellation *computo-symbolisme* : la sérialité des traitements et la localisation des symboles. Nous avons pu voir que l'approche connexionniste avait pu régler ce problème grâce à l'emploi de la dynamique : les traitements en parallèle et distribués sur l'ensemble de la cognition permettent de ne plus se fier uniquement à un traitement séquentiel et de rendre compte de la complexité de la cognition. L'interdépendance des composants dynamiques provoque des propriétés émergentes qu'on ne peut pas trouver dans une approche computationnelle. Il n'y a alors plus besoin de systèmes hiérarchisés puisque c'est l'auto-organisation du système dynamique qui permet l'émergence pertinente des différents résultats (Varela, 1999).

Bien que les théories connexionnistes permettent de résoudre certaines limites du cognitivisme, nous avons pu voir que l'approche dynamique des traitements distribués propre au connexionnisme ne permet pas de résoudre la question des représentations. La dynamique permet d'enrichir les conceptions de la forme et du contenu des représentations internes, mais elle ne mène pas à une cognition sans représentation (Clark, 1997). Le problème de l'approche *dynamisciste* est qu'elle se conçoit généralement comme une fin plutôt que comme une méthode d'étude théorique. En effet, la dynamique ne porte pas en elle les germes d'une conception complète de la cognition et il est nécessaire pour l'utiliser de faire appel à des théories externes pour lui donner une forme cohérente. Les chercheurs dynamiscistes ne peuvent prétendre n'avoir recours qu'à la dynamique pour expliquer la cognition sans faire preuve de mauvaise foi : ils s'appuient nécessairement sur des théories que la dynamique ne prend pas en compte, comme la distinction entre processus et contenu, ou la nature représentationnelle du contenu mnésique et cognitif. Généralement,

la dynamique sert de mise à jour d'une vision computationnelle et représentationnelle, tandis qu'elle pourrait être plus rigoureuse et être utilisée pour la modélisation de conceptions éenactives. Il ne faut pas perdre de vue que l'éenaction utilise la dynamique comme un rapport de nécessité, en aucun cas elle ne s'appuie sur elle car l'éenaction s'appuie sur l'autopoïese. L'éenaction ne part pas de la dynamique, elle y converge naturellement (la dynamique offre des outils de travail adaptés à la conception éenactiviste) mais l'éenaction part d'une conception non-représentationnaliste des interactions incorporées.

Traditionnellement la mémoire humaine a été modélisée de manière cognitiviste en un ensemble de systèmes dédiés au stockage de connaissances spécifiques selon leurs natures, encodées de façon abstraite et sérielle et dont le stockage est effectué en parallèle (voir le modèle SPI de Tulving, 1995). L'accès aux connaissances, dans ce type de modélisation, est le produit de l'activation d'une représentation mnésique et de la diffusion de cette activation à l'ensemble des informations associées en mémoire indépendamment du contexte de récupération (Anderson & Pirolli, 1984 ; Onifer & Swinney, 1981 ; Neely, 1977 ; Swinney, 1979).

Cette modélisation de la mémoire humaine a été remise en cause par une modélisation plus dynamique — dite « système unique » — qui considère que les connaissances ne sont pas stockées dans des registres particuliers qu'il faudrait activer, mais qu'elles émergent de la combinaison particulière de traces épisodiques, celle liée à l'expérience présente et celles liées aux expériences passées. La mémoire ici est appréhendée comme un système unique codant les expériences de nos interactions avec l'environnement (i.e., traces épisodiques) de façon multi-sensorielle et distribuée sur l'ensemble du cerveau (Brooks, 1978 ; Hintzman, 1986, 1988 ; Whittlesea, 1987 ; Versace et al., 2014). Dans cette modélisation l'accès aux connaissances est le résultat d'un couplage entre la situation de récupération et l'ensemble des traces épisodiques activées, en fonction de leurs similarités aux caractéristiques de l'épisode de récupération. C'est en ce sens que l'on considère que la connaissance émerge de la combinaison particulière des traces épisodiques sous l'effet de l'épisode de récupération. Tulving (1983) a théorisé ce concept sous le nom d'*ecphorie synergétique*, mais ses convictions cognitivistes l'ont limité à ne l'appliquer qu'à la manière dont le contenu mnésique est traité.

Les capacités très évoluées des systèmes complexes dans la cognition sont déjà largement utilisées (e.g., Busemeyer & Wang, 2015 ; Cretenet & Dru, 2009), et il semble impossible de s'en passer pour modéliser la mémoire de manière épistémologiquement cohérente. Labeye & Versace (2009) en ont dressé une revue explicite : pour que la mémoire ne rencontre pas de contradictions internes, elle se doit d'être épisodique (il n'y a de mémoire que par rapport à l'histoire d'un sujet), multimodale (chaque souvenir est lié à un contexte sensorimoteur et émotionnel), dynamique (la mémoire évolue constamment au cours de l'existence d'un individu, et ses souvenirs influencent son comportement)

et situationnelle (les connaissances émergent de la mémoire selon leur adéquation à la situation présente — car une réponse dépend toujours de sa question).

Utiliser l'écphorie synergétique comme garante de la dynamique mnésique dans un cadre énacliviste nous permettra de résoudre le problème de sérialité des processus et localisation des symboles propres au cognitivisme. Dès lors, nous baserons notre modèle sur une formalisation mathématique de l'écphorie synergétique proposée par Hintzman (1986, 1988) : MINERVA2. Cependant, un certain nombre de modifications vont être nécessaires pour que MINERVA2 puisse se revendiquer d'un énaclivisme radical.

1.2.2 Le problème du contenu mnésique

Nous avons pu voir dans notre présentation de l'énaclivisme qu'il n'y a pas de réponse simple à la question de la primauté de la signification. Qui est venu en premier, le monde ou l'image ? Rejeter la position cognitiviste de la « poule » comme quoi le monde est pré-signifiant — auquel cas nous ne faisons qu'interpréter plus ou moins fidèlement la signification de la réalité — ne mène pas forcément à un subjectivisme stérile. L'énaclivisme est un paradigme où le monde objectif n'est pas dissociable du monde subjectif. L'exemple de la couleur montre bien que la cognition ne traite pas les informations « vraies » du monde mais les interprète. Ce qui ne veut pas dire que l'existence du monde soit subjective, mais seulement que le monde ne peut pas objectivement être caractérisé en termes d'attributs. Il peut cependant l'être en termes de potentialité. Il existe des contraintes physiques, mais elles n'ont juste pas de signification. On ne peut pas abstraire objectivement le monde.

Selon Varela (1996, 1997) la structure d'un système est l'ensemble des relations concrètes qui ont lieu entre les éléments constitutifs du système. C'est-à-dire que seule la pertinence de l'expérience vécue du système est ce qui fait partie du réseau d'interactions système-environnement. La structure n'est pas antérieure au fonctionnement du système, c'est son application phénoménologiquement adaptée qui perdure et qui est ce par quoi nous le définissons ad hoc. La signification n'est pas antérieure à la modification d'un système, car ce en quoi consiste l'environnement n'est pas une cause de ce qui se produit mais une conséquence. L'idée de représentations cognitives présupposerait que la signification soit antérieure à l'activité cognitive. L'image est covariante avec le monde : cette conception largement constructiviste d'un système autonome est ce que Varela a appelé l'autopoïèse.

Si l'interaction entre l'individu et l'environnement est covariante, leur structure commune utilisée comme un tout en devient cohérente (Maturana & Varela, 1987). Plutôt que de chercher à essentialiser le monde et le système cognitif, c'est dans leur interaction phénoménologique que nous devons chercher l'information pertinente à utiliser pour modéliser la cognition : l'interaction entre le sujet et l'environnement comme source d'informations.

Selon Hutto & Myin (2012), la covariance est la seule forme d'information phénoménologique utile. En effet, si deux processus sont covariants ils sont probablement liés ; avec

un taux de probabilité variable, mais qu'on peut inférer en fonction de l'intensité de la covariance. La covariance est intimement liée à une fonction d'inférence, et elle permet d'extraire une information de corrélation sur leur origine commune (ou leur but commun). Dans ce cas, il n'est pas nécessaire de faire appel à ce que ces processus représentent (ni même s'ils se représentent l'un l'autre) : la seule chose importante est la probabilité que le traitement de l'un puisse provoquer le traitement de l'autre. Cette inférence de corrélation est parfaitement en accord avec la théorie de l'ecphorie synergétique où c'est la corrélation entre la situation présente et les traces passées qui permet d'inférer le souvenir. Le traitement généralement utilisé dans ce type de modélisation est la similarité entre la situation présente et les traces passées, mais la similarité est porteuse d'une forme de représentationnalisme (un contenu mnésique spécifique). Nous lui préférons la notion de covariance, porteuse d'une vision plus phénoménologique de la cognition (un traitement mnésique contextuel).

Bien entendu, l'information de corrélation portée par la covariance n'est pas un lien de causalité. Deux traitements covariants ne sont pas nécessairement liés. Mais dans une vision non cognitiviste de la cognition, la cognition n'a pas pour vocation d'explicitier une potentielle réalité objectivable. Dans une vision autopoïétique d'une cognition énonciviste, la corrélation est même plus efficace que la causalité en termes évolutionnistes. Pour optimiser la survie d'un organisme, ne vaut-il pas mieux voir un prédateur là où il n'y en a pas, plutôt que ne pas être totalement certain d'en voir un arriver ? La relation de corrélation est une interprétation, une inférence, qui une fois apprise à travers la covariance permet d'être réutilisée pour prédire des liens potentiels avec de nouvelles situations pour s'y adapter.

Pour que la cognition soit efficace, il n'est pas nécessaire de voir dans la covariance des éléments inter-représentatifs. Par exemple, les anneaux que l'on repère sur la souche d'un arbre covariant avec son âge. Mais les anneaux ne sont pas représentatifs de l'âge de l'arbre, il s'agit juste d'une information qui permet d'inférer son âge. Si la cognition ne peut pas traiter de représentations des éléments réels, elle peut cependant inférer un traitement à un autre juste en mémorisant les covariances rencontrées lors du passé du sujet. Un sujet cognitif peut alors associer un traitement à un autre et donner les mêmes résultats que ce que pourrait offrir une cognition représentationnelle : un comportement d'abstraction de concepts sans abstractions mémorisées.

En mémorisant des informations de covariance, si la cognition est covariante avec l'environnement cela signifie que l'environnement est en retour covariant avec la cognition. Dès lors, les traitements effectués en termes cognitifs seront toujours en adéquation avec les événements environnementaux : il ne sera jamais nécessaire de représenter l'environnement dans la cognition car les deux sont déjà intimement liés. L'information produite par la cognition à partir de l'environnement aura un sens dans cet environnement, elle n'aura même de sens *que* dans cet environnement et provoquera des difficultés d'adaptation

dans un autre environnement. Le monde-sujet est le système grâce auquel la cognition est possible, et ses traitements ne traitent que de ce système : il n'infère que lui-même (Hutto, 2005).

Dès lors, si l'information de covariance est suffisante pour rendre compte de la signification, un phénomène devient sa propre signification. Ce qui a deux conséquences : premièrement, les représentations ne sont pas nécessaires pour mettre au point des traitements de haut niveau (à condition que le niveau de complexité soit suffisant) ; et deuxièmement, une réunion processus-contenu devient nécessaire pour rendre compte des traitements de bas niveau.

Lorsque nous affirmons que le sujet est indissociable de son environnement, nous nous appuyons sur le fait que la relation sujet-environnement est une relation de co-dépendance et de co-définition. En effet, le sujet agit sur l'environnement de manière à modifier la façon dont celui-ci agit sur lui. La perception n'a de sens que comme lien définissant l'activité du sujet en fonction de son environnement, et bilatéralement l'action n'a de sens que comme lien définissant l'activité de l'environnement sur le sujet. Cette dynamique est un jeu de miroirs entre le sujet et son environnement, une adaptation aux interactions dynamiques où l'un se définit en fonction de l'activité de l'autre : c'est la dynamique d'une totalité monde-sujet en devenir (Barbaras, 1994). Pour le comprendre, nous pouvons prendre l'exemple du toucher sensoriel : le toucher est indissociable de la motricité de la main, sinon comment définir la dureté ou la rugosité ? Car l'expérience que nous faisons du monde n'a de lien qu'avec la manière dont nous interprétons cette expérience. Nous ne percevons pas la structure moléculaire d'un objet rugueux, de la même manière que nous ne l'inventons pas ex nihilo. Une expérience sensorielle est la manière dont nous interagissons avec cette même expérience, c'est notre expérience d'une motricité modulée par la perception (qui module la motricité en retour) qui est la seule information significative dont nous disposons. Cette information ne représente strictement rien, elle n'est que l'expérience momentanée d'une cohérence sensorimotrice. Nous ne pouvons pas remonter à l'origine de cette information pour trouver une hypothétique information objectivable, ce n'est pas une représentation. Cependant, c'est une information que l'on peut réutiliser dans une situation similaire. De plus, elle n'a pas besoin d'être exactement similaire, elle peut s'adapter par inférence et être au final plus adaptée qu'une représentation pré-définie.

Dès lors, la relation perception-action est autant co-définissante que la relation sujet-environnement : ce sont des boucles de dépendances et d'altérations réciproques (Merleau Ponty, 1945). Mais comme les différents traitements ayant lieu au cœur de la cognition ne sont pas plus différents que la perception et l'action, il nous faut aussi définir leur relation de co-dépendance comme une cohérence de covariance. L'information d'un traitement n'a aucun sens de manière indépendante des autres traitements, et pour modéliser une cognition éactive il va nous être nécessaire de n'utiliser que le traitement conjoint des

différentes composantes de manière récursive pour ne pas avoir recours à un niveau sub-symbolique qui serait contraire à une cognition énaïve.

On observe alors un renversement hiérarchique de la cognition : au lieu de considérer que le cerveau va être capable de traiter les informations sensorimotrices, on considère que les activités sensorimotrices vont permettre à la cognition d'émerger. Sur le plan évolutif, l'apparition de la cognition est concomitante avec l'apparition de la mobilité. Pour se mouvoir et se nourrir les animaux doivent mettre en lien des informations de la boucle perception-action, ils peuvent ainsi prévoir les comportements appropriés à la situation et s'adapter à leur environnement (Maturana & Varela, 1994).

De cette manière, la perception perd sa position traditionnelle d'input cognitif. En effet, la perception est un comportement sensorimoteur, elle consiste en une action guidée par la perception de cette action. Selon Merleau-Ponty (1942), « le comportement est la cause première de toutes les stimulations ». C'est un concept très bien imagé par l'expérience de Bach y Rita (1967) : les auteurs ont mis au point un dispositif sensitif relié à une caméra pour pouvoir traduire des informations lumineuses en stimulations tactiles (destiné aux aveugles). Mais les sujets n'ont été capables d'interpréter efficacement les stimulations tactiles que lorsqu'ils manipulaient eux-mêmes la position de la caméra. C'est en agissant sur leur perception du monde, que les sujets ont pu percevoir de manière efficace. L'information de covariance sensorimotrice n'est pas seulement suffisante pour rendre compte de la cognition, elle est également nécessaire.

Biologiquement parlant, les rétroactions et feedbacks sont globalement présents dans la quasi-totalité des structures neurologiques et sont nécessaires pour assurer le bon fonctionnement de la cognition (Varela et al., 1991). De même, les neurones miroirs sont un bon exemple de la non distinctivité entre perception et action puisque les mêmes patterns neurologiques s'activent pour une action perçue ou effectuée (Di Pellegrino et al., 1992). La covariance sensorimotrice peut théoriquement faire le lien entre le substrat biologique et le fonctionnement cognitif.

Les bases de la construction sensorimotrice de l'énaïve sont aussi présentes au cœur des travaux de Piaget. Ce que montre Piaget, c'est la façon dont les structures cognitives émergent de schèmes récurrents de l'activité sensorimotrice. Piaget ne considère pas que l'enfant doit acquérir des structures prédéfinies, mais que son apprentissage relève simplement d'une évolution sensorimotrice. On ne connaît un objet qu'en agissant sur lui, c'est donc de l'action et non de la perception qu'il faut partir pour comprendre en quoi consiste le développement de la connaissance, ce qui inclut aussi les pensées abstraites qui ne sont pas perçues mais produites (Piaget, 1970).

Le renversement des théories cognitivistes se retrouve régulièrement dans le constructivisme de Piaget. La distinction entre soi et l'environnement s'effectue à partir des interactions que l'enfant a avec son environnement, ce n'est pas la structure supposément

signifiante de l'environnement que l'enfant acquiert (Piaget, 1937). La construction du sensorimoteur est à l'origine de la capacité de l'enfant à penser son environnement et lui-même. Pour Piaget (1970), ce qui est important dans la connaissance, ce ne sont pas les actions prises isolément mais leur capacité à être adaptable à d'autres situations. Cette adaptation est une forme d'abstraction sensorimotrice à même de construire un concept sensorimoteur.

Voilà pourquoi si la cognition n'est pas représentationnelle, elle ne doit apprendre que les corrélations et covariances entre perception et action. S'il n'y a pas d'information essentialisée qui traverse les systèmes sensorimoteurs, le processus sensorimoteur doit être la mise en relation de la perception et de l'action sans distinction préétablie : pour le modéliser il va nous falloir un système qui traite le sensoriel et le moteur comme un tout ayant les mêmes particularités, les mêmes composantes sensorimotrices. MINERVA2 est intéressant pour cette raison : sonde et écho ont les mêmes composantes sensorimotrices (et non pas la perception pour la sonde et la motricité pour l'écho), il peut y avoir une production sensorielle au même titre qu'une production motrice (reste pour le modèle à apprendre la covariance entre la sonde et l'écho). Le système cognitif cherche à mettre en lien les informations sensorimotrices au moment présent en fonction de ce qu'il connaît déjà : le système mnésique n'est pas un module indépendant de la cognition, il est la capacité de réutiliser et d'adapter la manière dont nous traitons l'information.

Contrairement au symbolisme cognitiviste qui demande une distinction entre processus et contenu, la covariance énaïvistique doit être mémorisée sous la forme d'un processus pour pouvoir être réutilisable. Nous pouvons remarquer qu'au-delà de la limite représentationnelle du cognitivisme, la covariance sensorimotrice permet aussi de résoudre le problème de transduction récursive du représentationnalisme. La covariance sensorimotrice relie le sujet à son environnement, son auto-référencement phénoménologique relie une potentielle distinction entre symbole et sémantique, et le fait que le processus cognitif soit sa propre expérience sensorimotrice relie une distinction improbable entre conscience et expérience vécue. À travers la covariance sensorimotrice l'énaïvisme semble résoudre les problèmes rencontrés par le représentationnalisme cognitiviste.

Il n'existe à notre connaissance aucun modèle de mémoire n'apprenant que des covariances sensorimotrices comme le présupposerait l'emploi du paradigme énaïvistique.

1.2.3 Contextualisation et re-présentation

Maintenant que nous avons établi qu'une mémoire sans représentation doit nécessairement apprendre les covariances sensorimotrices, nous pouvons nous demander comment la cognition peut utiliser cette mémoire en situation. Lors du traitement d'une information, Varela (1993) oppose le traitement représentationnel au traitement contextuel. Comme nous avons pu le voir avec l'écphorie synergétique, le traitement mnésique dépend de la

situation présente (la réponse dépend de la question) : un souvenir est une construction qui permet à la mémoire d'interpréter les choses présentes en fonction de ce qu'on a pu apprendre dans le passé. Mais à la lumière de ces considérations, un problème majeur s'impose : s'il ne représente pas le passé, que signifie la nature passé du souvenir ? Les travaux de Sartre (1943) nous apportent des éléments de réponse : il ne faut pas considérer que le passé *était*, car avoir un passé n'est possible que pour un présent ; le passé *est* (pour un présent). Concrètement, cela signifie que les souvenirs ne représentent pas le passé, mais qu'ils sont le présent d'un individu. Quoi qu'il y ait en mémoire, les souvenirs ne sont pas un lien existant avec le passé : ils sont les conséquences des activités cognitives passées, mais ne sont utilisables qu'ici et maintenant. Nous reviendrons plus tard en détail sur les particularités temporelles de la cognition, mais retenons ici que le souvenir est une présentification des informations apprises : il est une contextualisation de la mémoire. Au sens de Havelange, Lenay & Stewart (2003) le souvenir, en tant que représentation en mémoire, est plus exactement une re-présentation : une activité de rendre présent. Rendre présent n'a rien à voir avec la récupération d'un souvenir, il s'agit fondamentalement d'une action, c'est un appariement : « L'image est un acte et non une chose » (Sartre, 1936).

Pour mieux comprendre comment concevoir la contextualisation d'un processus, prenons en exemple un cas en apparence extrêmement simple : la perception des couleurs. Le concept de couleur n'existe en tant que tel que dans sa relation avec un individu : sans organisme doté d'un système nerveux central et de dispositifs de couplage sensorimoteur particuliers, il n'y aurait que des ondes électromagnétiques. Mais la réception d'une onde électromagnétique n'est pas suffisante pour déterminer sa couleur car elle dépend de son contexte. La figure 1 (Kitaoka cité par Rasiwasia, 2017) nous présente une tarte aux fraises ayant des couleurs un peu sombres mais dont nous percevons clairement la couleur rouge des fruits. Pourtant les pixels utilisés dans cette image font tous partie d'un continuum chromatique allant du gris au vert. Contrairement aux apparences cette image ne renvoie aucune onde électromagnétique apparentée au rouge. La technique utilisée est assez simple et consiste à remplacer, dans une image à prédominance verte, des pixels gris là où la couleur rouge devrait être présente. Ces deux couleurs ont simplement des covariances inverses expérimentellement très marquées et nous construisons une perception rouge par contraste. Si la manière dont se construisent psychologiquement les covariances chromatiques dans l'histoire d'un individu n'est pas le sujet ici, nous pouvons remarquer que le traitement d'un pixel dépend du traitement de l'ensemble des autres, et que cette contextualisation se fait en fonction de la manière dont nous avons l'habitude de percevoir les couleurs : la mémoire permet le traitement au moment présent dans toute sa dimension dynamique. La perception de couleurs est une contextualisation dynamique des traitements présents en fonction des traitements passés.

A fortiori, et puisque la couleur n'existe pas en dehors de notre interaction avec les élé-



FIGURE 1 – « Red » strawberries par Kitaoka, 2017

ments chromatiques, c'est l'expérience vécue de la couleur qui permet de la définir (et non pas de la représenter). La mise en commun de toutes les informations passées et présentes est, selon la logique de l'ecphorie synergétique, une abstraction de concept. L'abstraction naît de l'interprétation mnésique, de l'utilisation de nos souvenirs ; elle ne flotte pas en mémoire de manière plus ou moins indéfinie (pour une revue, cf. Carbonnel, Charnallet & Moreaud, 2010). Loin de l'habituelle distinction entre représentations concrètes et représentations abstraites, la contextualisation énaïviste est toujours une forme d'abstraction (y compris pour la perception d'une couleur). Les différents niveaux de complexité dynamique permettent une abstraction plus ou moins poussée, mais utiliser une mémoire de covariance sensorimotrice par une présentification contextuelle se fait par abstraction continue. Cette contextualisation ne se retrouve pas seulement à un ensemble de couleurs, mais aussi à des ensembles multimodaux de couleurs, son, odeur, motricité, etc. C'est une démarche qu'on retrouve en psychologie du développement : Piaget considère que vers trois ans l'enfant acquiert la maîtrise de l'activité symbolique. Et l'acquisition de la capacité d'abstraction est inséparable de la boucle sensorimotrice (Smith & Thelen, 1993). Par exemple, le handicap d'un enfant qui ne peut pas bien manipuler les objets ayant un rapport avec l'eau va freiner ses capacités d'abstraction de l'eau.

Les particularités physiques de la couleur (longueur d'onde, coefficient de réflexion, etc.) sont généralement utilisées en IA comme prédéterminants que le système artificiel doit interpréter justement (c'est un constat qui vaut aussi dans un réseau connexionniste,

c'est aussi une vision totalement représentationnaliste). Pour l'événement, la détermination de ce qu'est visuellement un objet et de ses caractéristiques physiques est un processus complexe et continu qui provient d'un processus coopératif dynamique comportant un dialogue actif entre toutes les composantes perçues (visuelles ou autre, Chu, 2007 ; Ferrier et al, 2009).

Même si les modèles système unique — contrairement aux modèles multi-systèmes — mettent en avant le caractère émergent de la connaissance, leur conférant ainsi une certaine historicité (i.e., la façon dont le système s'adapte à la situation), ils considèrent que ce qui donne à une connaissance émergente son caractère de connaissance, c'est-à-dire l'expérience de mémoire, c'est la possibilité de coupler la trace de l'expérience présente aux traces des expériences passées. Or, comme le propose le modèle SCAPE (LeBoe-McGowan & Whittlesea, 2013 ; Whittlesea, 1997, 2002 ; Whittlesea & LeBoe, 2000, 2003 ; Whittlesea & Williams, 2000, 2001), une connaissance émergente n'est pas uniquement le reflet de l'expérience mnésique construite à partir de la combinaison particulière des traces épisodiques sous l'effet de l'épisode de récupération, elle est aussi la conséquence de la façon dont a été subjectivement évaluée l'accès à cette expérience mnésique. Ainsi, l'objet du processus d'évaluation est le processus de construction de l'événement mental lui-même, plus précisément la fluence subjective associée au processus de construction (i.e., comment se déroule le processus de construction). Par conséquent, c'est par un mécanisme d'inférence que l'événement mental construit ici et maintenant acquiert la vraisemblance d'un souvenir.

Le souvenir est construit ici et maintenant, mais pour pouvoir attribuer à cette activité le caractère de souvenir, il faut que la présentification de l'activité mnésique soit jugée comme passée. Ce jugement repose sur une perception subjective : la fluence. La fluence (voir Schwartz et al. 1997 ; Whittlesea, 1993 ; Oppenheimer, 2008) équivaut à une évaluation de l'activité de la mémoire face à une situation en particulier. Le caractère souvenir d'une activité mnésique est défini par la fluidité du traitement en cours. Donc, pour pouvoir la modéliser, il nous faut quantifier la fluidité avec laquelle l'ensemble des traces mnésiques se confronte à la situation sensorimotrice. Mais cette quantification devant rendre compte de la façon dont la situation sensorimotrice est subjectivement interprétée, elle ne peut pas se faire à l'aide d'un simple recueil des activités des traces mnésiques (comme le fait l'Intensité dans MINERVA2), sinon nous perdrons la relativité subjective de l'évaluation dans la production mnésique. En effet, la fluence va permettre de déterminer à quel point l'interprétation sensorimotrice obtenue semble être un souvenir. Il faut donc inférer la vraisemblance du souvenir en fonction de la situation actuelle. C'est pourquoi la fluence ne peut se modéliser efficacement qu'à partir d'un processus d'inférence (Schwartz, Benjamin & Bjork, 1997 ; Benjamin, Bjork & Hirshman, 1998).

Une mémoire basée sur le concept d'inférence est capable d'extrapoler un souvenir à

partir de très peu d'informations (Tenenbaum et al., 2011) en réduisant efficacement le bruit mnésique (Bitzer et al., 2014; Osth & Dennis, 2015). Elle est, certes, moins fiable dans une tâche précise qu'une mémoire cognitive mais est capable de s'adapter plus efficacement aux nouvelles situations. C'est d'ailleurs un des manques majeurs du modèle MINERVA2 pour pouvoir déterminer de manière pertinente le sentiment de reconnaissance (Benjamin, Bjork & Hirshman, 1998). Afin de modéliser une émergence d'abstraction sans représentation, nous tâcherons de permettre la contextualisation des covariances sensorimotrices par inférence. Cette méthode nous permettra de ne pas reproduire l'erreur connexionniste d'un sub-symbolisme pré-dynamique.

1.2.4 Meta-contextualisation de covariances fractales

Après avoir déterminé le rapport au monde de la cognition à travers le paradigme de l'ecphorie synergétique, puis en ayant déterminé ce qu'une mémoire sans représentation se doit d'apprendre (covariances sensorimotrices) ainsi que la manière dont elle doit être utilisée (contextualisation par inférence), nous pourrions penser avoir toutes les cartes en main pour modéliser une mémoire énonciviste. Cependant, ce serait retomber dans le piège de la sérialité connexionniste que de concevoir les covariances sensorimotrices comme l'input de la mémoire et la contextualisation abstraite comme son output sans prendre en compte l'interaction que la mémoire a avec son propre fonctionnement. Si nous réunissons le sujet et son environnement et si la mémoire n'apprend que des processus de covariance, ne faudrait-il pas également réunir les différents processus de covariances entre eux selon leurs propres covariances? De quoi traitent les processus de covariances s'il ne s'agit pas de patterns d'activation représentationnels?

Si nous nous limitons à une étude de la covariance entre le sujet et l'environnement et à une mise en contexte des traitements de covariances passés et présents, nous nous retrouverions à nouveau avec un problème de transduction entre les différents traitements de covariance. Mais les traitements de covariances ne sont pas plus distincts les uns des autres qu'ils ne le sont de l'environnement : il faut persister dans l'étude phénoménologique des traitements de covariance. Si le cerveau est capable de traiter des covariances environnementales et que la cognition n'est pas extérieure à l'environnement, alors les traitements de ces covariances environnementales font partie de l'environnement, donc de leurs propres traitements. La covariance sensorimotrice n'est pas un contenu au sens symbolique du terme, mais un processus (de la même manière que la réception chromatique ou la motricité sont des processus). Dès lors, nous ne pouvons pas comparer plusieurs covariances sensorimotrices comme des contenus symboliques, nous ne pouvons observer que leurs interactions : leurs informations de corrélation. De la même manière que la mémorisation d'une corrélation entre le sujet et l'environnement permet à l'énoncivisme de reproduire un schéma similaire dans une situation similaire sans avoir à faire appel

à une quelconque forme de représentation, la corrélation entre les différentes covariances apprises doit permettre une mise en relation adaptée dans une situation cognitive similaire. La cognition doit apprendre la manière dont elle traite l'information. Concevoir la cognition sans représentation demande l'utilisation d'un épiphénomène de covariances récursives.

Ces covariances récursives ne sont bien évidemment pas indépendantes : elles sont en constante interaction suivant le principe de contextualisation développé dans le chapitre précédent. Deux covariances sensorimotrices donnant lieu entre elles à une covariance apprise va les modifier en retour : de cette manière l'activation d'une covariance sensorimotrice pourra en activer d'autres en fonction de leurs corrélations respectives. Pour l'expliquer, nous pouvons concevoir un événement ponctuel de covariance sensorimotrice apprise comme une trace mnésique, et plus exactement une trace épisodique. Certes, nous ne concevons pas ici la trace épisodique au sens traditionnel du terme, puisque cette trace n'est pas une représentation de la situation apprise. Seule la manière dont l'information fut traitée est nécessaire pour pouvoir la reproduire. Dans la mesure où les covariances apprises de manière multimodale (contextuelle) sont apprises de manière ponctuelle et qu'elles jouent un rôle dans la production mnésique, nous pouvons considérer qu'une mémoire énaactive est fondamentalement épisodique.

Un des problèmes majeur rencontré avec les modèles à traces multiples existants est l'indépendance qui est généralement postulée entre les différentes traces. Chaque trace serait créée au moment où l'épisode mnésique est vécu et n'entreprendrait aucune interaction avec les autres traces. Pourtant, ces modèles utilisent toujours un ensemble de traces lors d'une récupération mnésique, on ne peut jamais récupérer une seule trace en particulier, c'est toujours une synergie. À quoi l'indépendance des traces pourrait-elle alors servir ? Probablement à simplifier la modélisation comme dans MINERVA2, mais ce n'est pas juste une simplification de forme, c'est une erreur méthodologique. Pourtant, Versace et al. (2009) postulent que la non-indépendance des traces n'est pas incompatible avec un système à traces multiples. En effet, une même zone neuronale participe au codage de nombreuses traces différentes : il y a une non-indépendance spatiale des traces. De plus, à chaque nouvelle situation les anciennes traces sont modifiées par la nouvelle, et cette dernière se construit en fonction des anciennes : il y a aussi une non-indépendance temporelle des traces.

À la base, le modèle Act-In de Versace et al. (2014) repose sur l'idée qu'une nouvelle situation ne va pas simplement réactiver l'ensemble des traces en lien avec celle-ci, mais aussi l'ensemble des traces en lien avec cet épisode de réactivation. Ce type d'activation en cascade demanderait par ailleurs un processus d'apprentissage en cascade pour pouvoir se souvenir de la totalité du traitement effectué. Nous retrouvons au cœur d'Act-In les mêmes prérequis nécessaires à la modélisation d'une mémoire énaactiviste. Premièrement,

les traces mnésiques reflètent toutes les composantes des expériences passées et sont distribuées sur de nombreux systèmes neuronaux (sensoriel, moteur, émotion et intention). Deuxièmement, la connaissance émerge du couplage entre la situation présente et les expériences passées, et ce en respectant la logique de l'écphorie synergétique. Troisièmement, le cerveau est une machine à catégoriser qui se développe en accumulant des expériences. C'est-à-dire que l'expérience de covariance sensorimotrice vécue est au cœur du phénomène de mémoire par apprentissage de corrélation. Enfin, l'émergence des connaissances requiert des mécanismes simples prenant part à la fois à l'apprentissage et à la récupération : toute activité mnésique est apprise comme le présupposerait la réunion entre processus et contenu. Les deux mécanismes mnésiques d'Act-In (activation et intégration) ne sont pas indépendants et coagissent en interaction.

Le modèle Act-In propose de considérer que le couplage contextuel entre composants sensoriels et moteurs ne s'opère pas uniquement entre les traces mais aussi à l'intérieur des traces. Ainsi, le mécanisme d'activation initié par l'épisode actuel activerait les composants sensorimoteurs des traces en mémoire similaires à ceux de l'épisode actuel (i.e., activation intra-composante : activation de différentes traces sur la base d'une même dimension). En parallèle, seraient également activés les autres composantes constitutives de la trace (i.e., activation inter-composantes : activation des différentes dimensions au sein d'une même trace). Enfin, un mécanisme d'intégration permettrait de coupler les différentes dimensions sensorimotrices pour former la trace. Par conséquent, dans Act-In, les traces ne peuvent pas être dissociées des mécanismes qui leur ont donné naissance, qu'il s'agisse d'une trace spécifique ou de l'ensemble des traces. L'activité mnésique d'activation inter-traces couplée à l'intégration multi-composantes permet de rendre compte simultanément de l'émergence de la connaissance et la modification du contenu de la mémoire : l'apprentissage et la reconnaissance sont issus du même processus.

Quelle que soit la situation, la nature de la connaissance émergente dépend de deux mécanismes en interaction : l'activation (inter-traces et inter-composantes) et l'intégration multi-modale. L'activation repose sur trois principes : la similarité entre l'indice présent et les traces mnésiques (comme le fait déjà MINERVA2), l'importance de l'association entre les composantes d'une trace (une contextualisation inter-composantes au cœur de chaque trace, ce qu'on retrouve déjà au niveau de la mémoire dans sa globalité), et la similarité entre les différentes traces (pour pouvoir produire des covariances en cascade à travers le processus d'intégration). L'intégration, quant à elle, est un phénomène permettant la diffusion d'activation de manière similaire à un processus d'inférences en cascade. C'est un processus dynamique, non-linéaire et parallèle à l'activation.

La cognition est fondamentalement — et par défaut — une machine à catégoriser, à inférer des corrélations. De plus, distinguer l'activation de l'intégration n'est pas tant une dissociation fonctionnelle qu'un élargissement du même processus local à une échelle plus globale. Ce qu'on observe au niveau neuronal : la plupart des activités se font au niveau

local, mais les connections entre réseaux locaux permettent la diffusion inter-composantes. De cette manière, l'abstraction (catégorisation) intra-composante est forcément plus importante que l'abstraction inter-composantes, et on retrouve la distinction de comportement au niveau local et global dans une distinction de processus complémentaires entre activation et intégration. Il n'y a pas de différence fondamentale entre activation et intégration, si ce n'est une question d'espace (réseau local vs réseau global) et de temps (le réseau local est par définition plus rapide que le réseau global).

L'apport principal que nous retiendrons du modèle Act-In est que l'interdépendance des traces et des composantes peut se traduire en invariance d'échelle. Les covariances récursives dont nous avons parlé au début de ce chapitre peuvent être considérées comme des processus dynamiques qui évoluent au cours du temps, selon les situations rencontrées. De cette manière, chaque covariance sensorimotrice apprise se comporte comme la mémoire dans sa globalité : elle apprend continuellement en fonction de la manière dont elle permet de traiter l'information. Act-In pourrait a priori se revendiquer être un modèle de mémoire sans représentation, mais dans la mesure où ses théories n'ont pas encore fait l'objet d'une formalisation mathématique, ce modèle ne peut pas être mis à l'épreuve dans des protocoles expérimentaux pour s'en assurer. Finalement, étant donné qu'une mémoire énoncive demande l'utilisation combinée de processus traitant des processus, il est un objet mathématique capable de s'adapter à nos besoins d'invariance d'échelle : la figure fractale.

Dans les sciences cognitives, la figure fractale est encore peu étudiée (probablement de par la difficulté de l'inscrire dans un paradigme cognitif traditionnel) mais se trouve être particulièrement efficace pour rendre compte de nombreux phénomènes cognitifs (voir Werner, 2010 ; Mattei, 2014). Globalement, et en accord avec sa nature d'adaptation généralisée, on peut retrouver les mêmes processus de récursion hiérarchiques dans la prédiction musicale comme la prédiction visuelle (Martins et al. 2017). La récursion fractale semble être une faculté non-spécifique de la cognition, nous l'utiliserons dès lors de manière indifférenciée par construction de corrélations.

Pour illustrer ce fonctionnement fractal, nous pouvons prendre une image simplifiée de l'utilisation contextuelle de réseaux de neurones dynamiques. En effet, dans un réseau de neurones le chemin emprunté par les informations dépend de la situation qui est traitée par ce réseau, mais il dépend aussi des chemins empruntés précédemment. Ces chemins précédents sont propres à un ensemble de situations précédentes qui ont employés ces neurones d'une manière qui leur était propre et qui ont modifié leur manière de traiter l'information (par la plasticité synaptique). Si nous définissons la modélisation énoncive par un réseau de neurones, les traces mnésiques qui le composent se retrouvent dans des chemins inscrits dans le réseau par une utilisation contextuelle d'une situation passée. En s'arrêtant ici, nous pourrions nommer ces chemins des représentations mnésiques.

Cependant, et dans la mesure où nous considérons que les traitements passés ne sont pas inscrits de manière statique dans la mémoire, mais continuent d'évoluer en fonction de l'activité de la mémoire, les traces mnésiques se comportent comme la mémoire dans sa globalité. En effet, chacun des chemins potentiels qui vont influencer la manière dont le réseau va transmettre l'information va avoir un effet différent selon l'information à traiter et ces changements d'activité vont eux aussi être mémorisés. Nous pouvons apercevoir ici le besoin de l'utilisation fractale des processus qui produiraient ces chemins : chaque chemin potentiel va influencer différemment le chemin parcouru en fonction à la fois de la situation à traiter, mais aussi en fonction de ses influences passées. Chacune des possibilités de chemin est constamment influencée par sa propre activité au cours du temps, nous retrouvons dès lors le même type de fonctionnement pour définir à la fois la potentialité d'un chemin, et le chemin qui sera bel et bien emprunté. L'utilisation fractale de la mémoire prend tout son sens dans la contextualisation de l'information à traiter comme un tout.

Au final, si le contenu de la mémoire est une activité, il est nécessaire de réunir processus et contenu dans une conception fractale de la mémoire. Nous avons jusqu'ici parlé de la mémoire comme d'un système de contextualisation de covariances sensorimotrices, nous pouvons désormais parler de meta-contextualisation de covariances fractales. La meta-contextualisation devrait renforcer les capacités abstraitives d'une mémoire énaïviste : c'est-à-dire non seulement permettre la construction, l'apprentissage et la récupération d'abstractions (comme nous avons pu le voir au chapitre précédent), mais aussi manipuler ces abstractions par émergences en cascades (e.g., la mise en relation syllogique). La meta-contextualisation devrait permettre de retrouver l'interactivation relative des traces développée par Act-In : l'intensité d'activation d'une trace dépendra de ses co-dépendances relatives avec les autres traces.

1.2.5 Mémoire et construction temporelle

Les bases théoriques pour la modélisation d'une mémoire énaïviste semblent réunies, mais il reste une critique du connexionnisme sur laquelle nous ne nous sommes pas encore penchés : la question temporelle. Le temps est fondamentalement lié à la mémoire : non seulement c'est au cours du temps que se crée la mémoire, mais c'est aussi par la mémoire que l'on se crée une représentation du temps. Lorsque l'on se base sur des conceptions dynamiques, un modèle de mémoire ne pourra jamais être complet sans prendre totalement en compte l'aspect temporel de la mémoire. L'approche connexionniste considère que le temps est un absolu sur lequel se construit la cognition, pourtant non seulement le temps ne se déplace pas de manière linéaire (il est une interprétation mnésique, comme nous allons le voir), mais surtout nous ne mémorisons pas de représentations épisodiques sur lesquelles baser cette linéarité. Nous avons pu voir dans le chapitre traitant de la contex-

tualisation qu'un souvenir — sous l'égide énéactiviste — est distribué sur l'ensemble des traces mnésiques. Ce qui signifie qu'un épisode mnésique ne peut pas être récupéré seul, les traces perdent toute distinctivité temporelle, et il peut sembler a priori impossible de construire un comportement mnésique temporel à partir d'une mémoire ne mémorisant pas le temps comme un absolu. Pourtant l'étude de la question du temps est pour Varela (2002) une question de neurophénoménologie, ou plus explicitement une expérience de la dynamique neurologique en fonction de l'environnement. Après avoir défini ce qu'est le temps pour la mémoire, nous verrons que la contextualisation fractale permet l'émergence d'un tel comportement temporel.

La conception intuitive que l'on a du temps ne correspond pas aux avancées scientifiques de Physique Théorique développées depuis le XX^e siècle. Premièrement, nous avons tendance à penser le temps en observant l'évolution d'objets au cours du temps. Pourtant, le temps n'est pas une dimension indépendante de l'espace (l'espace et le temps sont fondamentalement liés, comme les trois dimensions spatiales sont liées), on ne peut pas se limiter à étudier l'espace dans le temps comme si celui-ci était une référence absolue et immuable (Einstein & Infeld, 1948). Deuxièmement, et même si nous ne semblons pas avoir d'emprise sur le temps, celui-ci est soumis à nos propres caractéristiques physiques : le temps dépend de la vitesse de l'observant, notre relation corporelle avec celui-ci est dynamique (Hawking, 1993). Troisièmement, une des principales caractéristiques intuitives du temps serait son avancée inexorable. Pourtant, le temps ne se crée pas en même temps que nous l'observons, il est déjà totalité : comme l'espace existe quelque part en tous points, le temps existe déjà et encore (Greene, 2005). Enfin, pour résumer, le temps n'est pas ce qu'il semble être. Il ne s'écoule pas simplement dans une seule direction, et le futur existe simultanément avec le passé : « La distinction entre passé, présent et futur n'est qu'une illusion, même si elle est tenace » (lettre d'Einstein à Besso cité par Prigogine & Stengers, 1979). Le modèle SAM (Raaijmakers & Shiffrin, 1981) par exemple, considère que le temps est une composante absolue sur quoi se construit la mémoire, ce qui exclue toute forme d'interprétation temporelle, et pose même certains problèmes dans la conception scientifique du temps.

Mais dans ce cas, comment se fait-il que notre vision intuitive du temps soit tellement éloignée de ses caractéristiques physiques ? Pour le comprendre, reprenons l'analogie développée par Klein (2009) sur le passager d'un train : en étant assis côté fenêtre, le nez collé contre la vitre, on a l'impression que le paysage défile, alors que c'est le train — donc soi — qui est en mouvement. La raison en est simple : la notion de mouvement est relative, elle dépend d'un référentiel. De même, nous avons l'impression que le temps défile devant nous car nous nous considérons comme conceptuellement stable. Pourtant une multitude de changements physiques et psychologiques s'opèrent constamment en nous, nous ne sommes plus la même personne au cours du temps qui passe ; mais dans un souci de cohé-

rence psychologique nécessaire, nous en sommes persuadés. C'est ce qu'on appelle l'ipséité, il s'agit en quelque sorte du penchant psychologique de ce que Varela (1974) appelait l'autopoïese. Le développement d'un système autonome tel que décrit par Varela ne signifie pas que le système est isolé, bien au contraire il décrit le mécanisme de conservation d'une logique propre au système malgré toutes les interactions qu'il a constamment avec son environnement (et qui le modifient continuellement). La construction temporelle détermine ce par quoi l'interprétation sensorimotrice est possible : le système en lui-même. À travers l'ipséité la construction perceptive peut inférer des corrélations menant à une notion d'intentionnalité et de « soi » (c'est de cette manière que nous construisons une constance sous la forme d'une personnalité, laquelle permet la construction de la conscience, Sartre, 1943). Bien entendu, la notion de « soi » n'est pas le sujet de ce travail, mais pour ne pas reproduire l'erreur cognitiviste consistant à reléguer à plus tard la question de la conscience au risque de développer des contre-sens (rappelons ici que pour le cognitivisme la réflexion est possible sans conscience de soi), nous nous devons de préparer le terrain de manière cohérente avec la phénoménologie de la conscience. Pour que la mémoire soit cohérente avec le paradigme énaïvistique radical, elle ne doit pas contenir de représentations épisodiques mais doit être capable de construire un comportement temporel émergent par cohérence de contextualisation. En définitif, nous ne percevons pas le temps en lui-même, mais l'évolution des objets avec lesquels nous interagissons. La même chose est vraie avec les souvenirs, on détermine leur position temporelle en fonction de la manière dont nous nous en souvenons, en fonction de l'ensemble des traces mnésiques. C'est fondamentalement l'aspect actif de la perception qui donne à la temporalité ses racines dans un individu vivant (Varela, 2002). Il s'agit de l'évolution d'une contextualisation de covariances, et la mémoire énaïvique est prévue pour être capable d'en rendre compte.

L'évolution de contextualisation de covariances étant relativement complexe à saisir, prenons en exemple la tâche de Fisher (1967) pour l'expliquer en terme de traces mnésiques. Cette tâche consiste en la présentation d'une suite d'images dessinées au trait noir. La première image représente le visage d'un homme, tandis que la dernière représente le corps d'une femme. Chaque image les séparant consiste en une légère modification de l'image précédente amenant à l'image suivante, et la suite donne lieu à un dégradé où il n'y a pas de frontière nette entre la première et la dernière image. Lorsqu'on présente cette suite d'images à un individu, la limite entre la première image et la dernière dépend de l'ordre de présentation. Globalement la limite est définie par les sujets comme étant plus proche de la dernière image que de la première. Cet effet est dû à une forme d'habitation du traitement où la manière dont les images sont traitées va être déterminante pour la manière dont les sujets traitent les images suivantes. L'apprentissage de la manière dont on traite l'information va entraîner une certaine forme de prédiction perceptive par cohérence du traitement global de la tâche. C'est une situation qu'on retrouve quelle que

soit la situation interprétée, par exemple au moment présent vous êtes en train de lire et pourtant vous prédisiez la suite de ce que vous lisez tout en interprétant en fonction de ce que vous avez déjà lu au passé (en effet, tout en comprenant son sens, vous seriez surpris si cette phrase ne se terminait pas comme vous l'aviez poney). L'interprétation induit une recherche de cohérence temporelle.

En termes de traces mnésiques, chaque nouveau traitement s'effectuera en fonction des traitements précédents, il suffit donc d'apprendre la manière dont on traite l'information pour permettre l'émergence d'un phénomène de primauté. Les premières traces étant plus présentes en mémoire que les dernières, celles-ci joueront un rôle plus élevé dans l'émergence d'un souvenir que ces dernières. Les images proches de la fin restent fortement corrélées à la première image puisqu'elle a déterminé comment interpréter les suivantes (et les suivantes s'influencent entre elles également). Cependant cette seule explication n'expliquerait pas pourquoi la limite ne se trouve pas toujours à la dernière image pour les sujets. Le fait est que lorsqu'une nouvelle covariance est apprise, celle-ci modifie la totalité des traces existantes en fonction de leur manière d'interpréter la situation. Apprendre une nouvelle situation modifie la capacité des traces précédentes à interpréter une situation et nous retrouvons un effet de récence où les dernières informations interprétées restent très présentes en mémoire. Au lieu d'un renforcement continu des premières traces, nous nous retrouvons avec des évolutions relatives aux changements rencontrés dans les contextualisations de covariances.

En ce sens, MINERVA2 aurait le potentiel d'acquiescer un comportement temporel. Il est en effet capable d'assimiler de nouveaux souvenirs de manière chronologique sans qu'aucune distinction de traitements ne soit faite lors d'une re-présentation. Cependant, puisque ces traces n'évoluent pas, il ne construit pas de cohérence temporelle. Dès lors, d'autres limites ontologiques en découlent : ses souvenirs sont indépendants et déterministes, de plus il ne peut faire l'objet ni de comportement éactique, ni de méta-abstraction. Toutefois, MINERVA2 a été sujet à une tentative d'amélioration temporelle : MINERVA2 Étendu (Damas, Mille & Versace, 2002). Les auteurs ont proposé une amélioration associant la création de chaque nouveau souvenir à l'ensemble des anciens souvenirs, ce qui permet une construction temporelle cohérente. Cependant, ce modèle a deux grandes insuffisances. Premièrement, la mémoire à court terme est provoquée : elle n'émerge pas du fonctionnement du modèle mais est assimilable à un module d'altération mnésique, ce qui nous éloigne d'une conception unique de la mémoire. Deuxièmement, la dynamique cognitive est unidirectionnelle : les souvenirs précédents influencent les nouveaux, mais un nouveau souvenir n'influence pas les anciens, ce qui compromet la validité dynamique du modèle.

La contextualisation fractale que nous avons présentée au chapitre précédent permettrait l'évolution des traces mnésiques comme autant de systèmes éactifs en interaction

perpétuelle. Cela va permettre de développer l'émergence d'une cohérence temporelle à travers les phénomènes de primauté et de récence (pourtant généralement présentés comme la preuve d'une dissociation entre mémoire à court terme et mémoire à long terme). L'énactivisme radical permet à la mémoire de développer des liens de corrélations entre les différentes traces en suivant une logique temporelle car il s'agit de la corrélation la plus simple à exploiter de manière cohérente. Il est question du même type d'émergence transitive que nous avons pu voir dans le chapitre précédent, mais si la méta-contextualisation de covariances fractales permet bel et bien de reproduire les caractéristiques intra et inter composantes de Act-In, sa modélisation devrait de facto permettre l'émergence d'un comportement temporel cohérent.

1.3 Conclusion et problématique

Rejeter le postulat de l'existence de représentations en mémoire nous a mené vers les apports de l'énactivisme radical. Les cinq critiques soulevées des paradigmes cognitivistes et connexionnistes ont ici été résolues par l'emploi d'une méthodologie phénoménologique propre à l'énactivisme radical.

Premièrement, pour résoudre le problème de la sérialité des processus et la localisation des symboles propre au computo-symbolisme, nous emploierons la modélisation du phénomène d'ecphorie synergétique dans une mémoire à système unique. L'ecphorie synergétique n'est pas suffisante pour rendre compte d'une mémoire sans représentation, mais il s'agit d'une base solide pour pouvoir mettre en place un processus éactif radical. Pour ce faire, nous nous baserons sur MINERVA2 qui propose déjà une formalisation mathématique du processus d'ecphorie synergétique (Rousset, 2000).

Deuxièmement, en rejetant l'idée que la mémoire puisse contenir des représentations, nous nous sommes confrontés au « hard problem of content » : quelles informations contient la mémoire ? L'utilisation de la covariance sensorimotrice est non seulement suffisante pour se passer de représentations, mais aussi nécessaire pour obtenir une perception efficace, une interprétation pertinente de l'environnement. L'utilisation de la covariance sensorimotrice permet alors de résoudre le problème de transduction récursive. MINERVA2 en serait théoriquement capable puisqu'il réunit les mêmes composantes de sonde et d'écho, à condition de lui faire apprendre la covariance entre les deux au lieu de n'apprendre que le contenu de la sonde (la perception entrante).

Troisièmement, pour éviter toute forme de sub-symbolisme les covariances apprises doivent pouvoir être réutilisées de manière contextuelle et par inférence. Tout traitement mnésique devient ainsi un processus d'abstraction, et les capacités de mémorisation et de récupération d'abstractions devraient être possibles. Le calcul de l'intensité de MINERVA2 sera alors modifié pour rendre compte de la fluence relative du traitement de contextualisation par inférence.

Quatrièmement, dans la mesure où les covariances sensorimotrices apprises font aussi partie de l'environnement, et pour pouvoir les confronter de manière à respecter les contraintes de l'énactivisme radical, nous devons aussi les traiter en fonction de leurs covariances. Il est un modèle théoriquement capable d'opérer ce type de contextualisation récursive : le modèle Act-In. Il n'a malheureusement pas encore fait l'objet d'une formalisation mathématique, mais nous pouvons simuler son fonctionnement grâce à l'emploi d'une meta-contextualisation des covariances fractales. Appliquer ses théories dans un modèle de type MINERVA2 va dès lors nécessiter de réunir processus et contenu en s'assurant que l'apprentissage et la contextualisation de ses traces mnésiques se fassent de manière fractale.

Enfin, l'absence de représentations mémorisées ne signifie pas qu'un modèle énonciviste radical perde ses capacités temporelles. La meta-contextualisation de covariances fractales donne aux traces mnésiques la possibilité d'évoluer de manière cohérente en fonction des interactions qu'elles ont eues entre elles. Le quatrième point permet aussi de résoudre ce problème. Si l'interaction entre les traces mnésiques s'effectue de manière réellement cohérente, nous devrions, en plus des capacités abstractives, observer dans le modèle un comportement temporel également cohérent.

Modéliser une mémoire sans représentation va nous demander de formaliser mathématiquement les quatre déterminants développés précédemment : l'écphorie synergétique, la covariance sensorimotrice, l'inférence de contextualisation et la réunion processus-contenu sous forme de fractale.

Modélisation

2.1 Modéliser la mémoire

Le problème de la modélisation dans le cadre du paradigme cognitiviste est qu'on demande à un modèle de représenter le monde comme si le monde tel qu'on le connaît était objectivement le monde tel qu'il est. Ce faisant, un tel modèle ne serait pas un modèle de la cognition humaine car il occulterait la place de l'interprétation humaine (i.e., celle de la signification) dans la perception du monde. Ce problème méthodologique provient d'une prise de position ontologique. Le cognitivisme remet fondamentalement en question l'aspect subjectif et contextuel de la signification. Dès lors, se réclamer du cognitivisme tout en prenant en compte la dynamique contextuelle de la signification ne peut que générer un conflit épistémologique (généralement résolu par les cognitivistes en ignorant tout simplement le problème). Nous considérons que la prise en compte de ce problème ne nous éloigne pas de la science pour nous renvoyer dans le champ de la philosophie. C'est justement le fait d'ignorer ce problème qui contredit l'idée que ce problème est scientifique.

C'est pourquoi nous nous devons d'affiner notre manière de réfléchir à partir de ce que nous connaissons de la psychologie : si la psychologie démontre que tous les processus cognitifs sont contextuels et dynamiques, il faut que nous prenions en compte le fait que notre réflexion sur cette contextualité dynamique soit dynamiquement contextuelle. Ce doit être le cas pour la modélisation. La modélisation est par nature une simplification de phénomène qui vise à optimiser notre compréhension. La contextualisation cognitive n'est pas une représentation du monde tel qu'il est dans l'absolu, c'est une contextualisation de notre réflexion à la dynamique du monde. Si une cognition non-représentationnelle est capable de prendre en compte cette distinction meta à travers le paradigme de l'énaction, un modèle représentationnel ne peut pas représenter la réalité scientifique (cette idée n'a même aucun sens).

Pour l'imager, on peut réfléchir à la notion d'espèces animales. La science est capable de catégoriser, de délimiter les propriétés d'une espèce vis à vis d'une autre. Mais est-ce que la notion d'espèces différentes est une représentation de la réalité du monde qui contient des espèces objectivement différentes ? Ou bien est-ce que la notion d'espèce fait plutôt référence aux différences d'interactions qu'on est capable d'avoir avec celles-ci ? (l'étude scientifique est aussi une interaction de l'Homme à son environnement) Dans les

faits, si un individu de l'espèce humaine est différent d'une cellule eucaryote dont notre espèce est issue (dans le sens où l'interaction possible est différente), il n'y a jamais eu de « saut » d'une espèce à une autre. Nous sommes de la même espèce que nos parents (avec de légères mutations génétiques), qui sont eux aussi de la même espèce que les leurs et ainsi de suite jusqu'à notre ancêtre monocellulaire. La distinction entre les espèces n'est pas objective, ce qui ne veut pas dire qu'elle soit scientifiquement fausse, mais elle est relative à notre manière de penser la science, elle n'est pas une représentation des différences réelles entre les espèces, mais de ce qu'il est utile de différencier.

Pour répondre à la remarque possible que dans ce cas la science ne serait pas aussi efficace si elle n'était pas garante de la connaissance objective, il faut bien préciser que la science est une méthode eidétique. C'est-à-dire qu'elle traite de son propre sujet d'étude. Comme les mathématiques sont les méthodes qui permettent d'analyser les mathématiques. La science est la meilleure méthode connue pour obtenir des connaissances objectivables dans le cadre de la science. Différencier des espèces est utile dans le cadre scientifique même si cette différenciation n'est pas une représentation du monde, ce n'est qu'une re-présentation de notre interprétation de son utilité scientifique. De la même manière que les nombres complexes sont utiles dans le cadre des mathématiques, même si le nombre imaginaire i ne représente rien d'autre que l'utilisation qu'on en fait. De même, la difficulté d'expliquer clairement nos connaissances scientifiques en terme de mécanique quantique est liée au fait qu'elle soit contre-intuitive à notre expérience newtonienne du monde. L'intrication quantique n'est pas fausse pour autant, elle est juste dépositaire de l'expérience scientifique de nos méthodes scientifiques utilisées pour la démontrer.

Pour en revenir à notre propos, modéliser la cognition doit rendre compte de la manière dont la cognition se construit, évolue et s'adapte. Modéliser la cognition c'est donc modéliser la mémoire (« la mémoire est la forme même de la cognition », Tiberghien, 1997). La modélisation permet de s'assurer qu'une théorie est applicable de manière pertinente, ainsi que de confronter plusieurs théories pour observer comment elles interagissent entre elles. Elle permet surtout de confronter le comportement d'un modèle à celui de sujets humains par la simulation de protocoles expérimentaux (dans une tentative d'infirmer la capacité du modèle à reproduire le comportement humain). Réussir à reproduire le comportement mnésique d'un sujet humain n'est en aucun cas une preuve que la mémoire humaine fonctionne de la même manière, mais elle permet de montrer que les théories sous-jacentes du modèle sont capables d'expliquer un comportement mnésique. Nous allons ici mettre à l'épreuve modestement la théorie énaactive pour s'assurer de sa capacité à faire émerger des comportements cognitifs psychologiquement pertinents dans le cadre d'une mémoire sans représentation.

2.2 Le modèle MINERVA2

2.2.1 Présentation de MINERVA2

Puisque nous allons partir du modèle MINERVA2 (Hintzman, 1984, 1986, 1988) pour pouvoir développer une mémoire sans représentation, il est nécessaire de présenter rapidement ce en quoi consiste ce modèle et sa description d'une mémoire représentationnelle. Hintzman représente le contenu de la mémoire sous forme d'une matrice (voir figure 2 ci-dessous).

$i \backslash j$	1	2	3	...	m
1	1	1	-1	...	-1
2	0	0	1	...	1
3	1	0	-1	...	0
...
n	-1	-1	0	...	0

FIGURE 2 – Matrice inter-traces et inter-composantes

Chaque ligne de la matrice représente une trace ($i = 1...n$ — où n représente le nombre total de traces présentes en mémoire) et chaque colonne une composante de l'expérience codée par la trace (également appelé dimension de la trace : $j = 1...m$ — où m représente le nombre total de composantes partagés entre toutes les traces). Dans une situation donnée — représentée de la même manière que chacune des traces, par une configuration de composants — la nature des connaissances susceptibles d'émerger dépend de deux mécanismes, l'activation inter-traces et l'activation intra-trace. L'activation inter-traces désigne la diffusion d'activation vers et entre les différentes traces (verticalement dans la matrice) à partir de l'expérience présente.

Cette propagation n'est pas séquentielle mais parallèle, car elle intervient séparément et indépendamment au niveau de chacune des composantes et dépend essentiellement des similarités entre traces au niveau de leurs composants. Par exemple, la vue de l'image d'un chien active en mémoire toutes les traces correspondant à des situations dans lesquelles l'individu a été confronté à des stimuli visuellement similaires. Mais, bien entendu, la vue de ce chien doit aussi activer, au niveau de chaque trace, les autres propriétés associées

à la dimension visuelle activée, par exemple des propriétés sonores, olfactives, tactiles, etc. Cette activation au sein de chaque trace entre les dimensions qui la composent correspond à l'activation intra-trace. L'activation intra-trace est donc une activation inter-dimensionnelle, les dimensions pouvant être des composantes sensorielles différentes, mais aussi des propriétés différentes au sein d'une même composante (e.g., forme, couleur, etc. pour la composante visuelle). Ainsi ces deux mécanismes interviennent simultanément et s'influencent mutuellement. Ils sont le résultat du couplage entre l'expérience présente et les traces d'expériences passées. L'activation inter-traces se diffuse à partir des propriétés directement activées par l'expérience présente (donc perceptivement présentes), mais aussi à partir des propriétés activées par l'intermédiaire de l'activation intra-trace.

MINERVA2 formalise mathématiquement la gestion d'informations mnésiques à partir d'éléments non-abstractifs. Pour ce faire, le modèle représente un souvenir épisodique par une suite de chiffres issus de l'ensemble $(-1, 0, 1)$, où chaque valeur correspond au niveau d'activation d'une composante. Par souci de clarté, nous appelons la matrice de mémoire X , et la dimension (ou composante) j de la trace i est appelée x_{ij} . Ces composantes représentent toutes une caractéristique primitive qui est soit activée (1), soit inhibée (-1), soit sans activité particulière (0) lors du souvenir. Nous avons donc au final un souvenir épisodique multimodal capable de rendre compte de la concordance d'activité entre toutes ces caractéristiques lors d'un épisode mnésique.

En plus du stockage d'informations, MINERVA2 est capable de récupérer un souvenir par un système analogue au sonar. En effet, un souvenir ne dépend pas seulement des informations contenues en mémoire, mais aussi du contexte actuel (une réponse dépend forcément de la question posée). Dans ce modèle, la question est représentée par une sonde structurée comme une trace : c'est un vecteur multimodal auquel est comparé l'ensemble des traces présentes en mémoire. Chaque ancien souvenir du modèle sera alors activé plus ou moins intensément selon sa similarité avec la sonde, ceux-ci vont alors s'associer et provoquer le souvenir propre au contexte actuel. De cet appariement global émerge alors un écho capable de représenter des récurrences statistiques : si seules quelques récurrences de la sonde sont activées en mémoire, le souvenir sera précis et donc concret ; au contraire, si de nombreuses récurrences sont activées le souvenir sera global et son abstraction représentera les caractéristiques propres à la création de catégorisations. Le modèle est ainsi capable de généraliser les expériences passées pour en faire une catégorie (sans avoir besoin de mémoriser cette catégorie puisqu'elle émerge lors du rappel).

C'est ainsi que le modèle MINERVA2 est capable de résoudre les problèmes de sérialité des processus et de localité des symboles, puisque les informations et le traitement est distribué sur l'ensemble des traces et dépend de la situation actuelle, comme le préconise l'emploi de l'ecphorie synergétique.

Pour ce faire, comme présenté dans la figure 3, lorsque la sonde est comparée aux traces

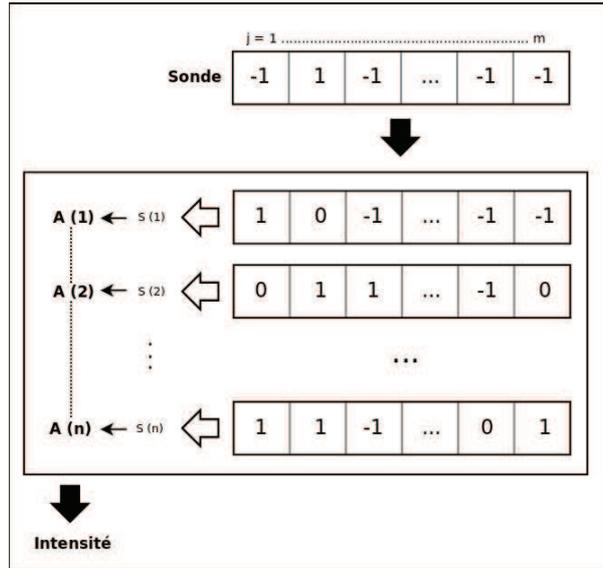


FIGURE 3 – MINERVA2 : Génération de l’intensité

existantes, une valeur de similarité est calculée pour chaque trace par un simple produit scalaire : $S_i = \frac{1}{M_r} \sum_j (\theta_j \cdot x_{ij})$ où θ_j représente la valeur j de la sonde, x_{ij} représente la valeur j de la trace i , et M_r représente le nombre de composantes significative (i.e., Les composantes dont le produit scalaire n’est pas issu de la multiplication de deux zéros). Le résultat obtenu est un flottant de l’intervalle $[-1; 1]$ reproduisant les valeurs d’activations et d’inhibitions des composantes, dont l’amplitude est amplifiée en le portant au cube tel que, pour l’activation d’une trace, $A_i = S_i^3$. La valeur d’activation obtenue permet dès lors de quantifier la prégnance des souvenirs selon leur similarité avec la sonde : ils pourront ainsi influencer le modèle plus ou moins intensément soit en terme d’activation brute (pour une valeur positive), soit en terme d’inhibition (pour une valeur négative). Il s’agit du mécanisme d’activation inter-traces.

Le modèle peut ainsi calculer une valeur d’intensité de son activité relative à la sonde grâce à la somme de ces activations. L’intensité de l’écho ainsi calculée ($I = \sum A_i$) représente en quelque sortes la familiarité de la sonde en fonction des souvenirs passés. Plus I a une valeur positive importante, plus la sonde sera semblable aux traces. Bilatéralement, plus I aura une valeur négative importante, plus la sonde sera différente des situations passées. Enfin, si I est proche de 0, aucune reconnaissance particulière n’émergera du modèle.

De même, comme présenté dans la figure 4, l’écho est une représentation de la totalité des activités combinées de chaque souvenir, et peut ainsi être calculé grâce à la moyenne des traces pondérées par leurs activations (ce qui permet de rendre compte de la complémentarité entre la diffusion inter-traces et intra-trace). En effet, le contenu de l’écho dépend de la synergie des traces en fonction de la sonde, donc les traces les plus activées par la sonde (représenté par l’intensité des valeurs d’activations — que ce soit de manière

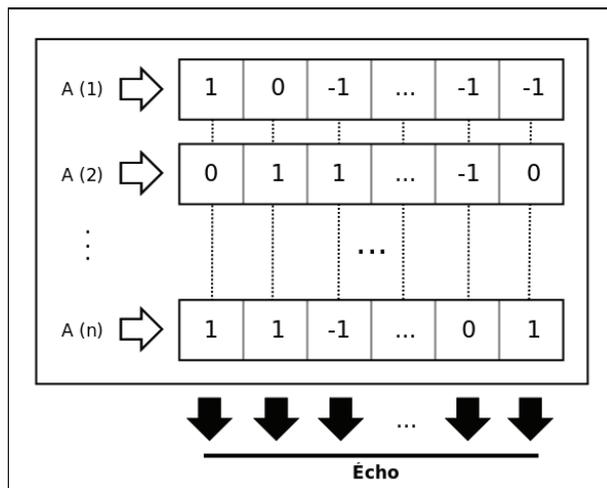


FIGURE 4 – MINERVA2 : Génération de l'écho

positive ou négative) sont celles qui vont influencer le plus fortement l'émergence de l'écho grâce à l'équation : $C_j = \sum(A_i \cdot x_{ij})$. C_j représente ici la composante j de l'écho, A_i représente l'activation de la trace i et x_{ij} la composante j de la trace i . Les valeurs de x_{ij} correspondant à une trace très activée influencera plus fortement le contenu de l'écho que les valeurs de x_{ij} ayant une activation de trace relativement faible. Ce qui signifie qu'une valeur x_{ij} différente de la sonde mais appartenant à un souvenir épisodique très activé sera capable d'influencer fortement l'émergence du souvenir ainsi reconstruit. L'écho sera alors conforme à l'ensemble des activités internes du modèle et représentera la réponse du modèle à la question de la sonde. Il s'agit du mécanisme d'activation intra-traces.

Enfin, la sonde est enregistrée comme nouvelle trace épisodique dans le pool mnésique. Mais pour modéliser le phénomène d'oubli, chaque composante de la sonde (θ_j) est encodée avec une probabilité L (définie arbitrairement selon les besoins de l'expérience). Ce qui signifie, par exemple, que pour un learning rate de 0.6 la composante j de la sonde a 60% de chances d'être encodée comme nouveau souvenir (et 40% de chances d'être enregistrée à 0). Ce nouveau souvenir sera désormais utilisé lors de la prochaine présentation d'une sonde en même temps que l'ensemble des autres traces, ce qui permet au modèle d'être influencé par ses expériences précédentes.

2.2.2 Modifications nécessaires pour modéliser une mémoire sans représentations

MINERVA2 est particulièrement efficace dans les phénomènes de catégorisation, et son système de question/réponse lui permet de rendre compte de phénomènes mnésiques dynamiques (comme le rappel indicé). Ce modèle possède en outre une simplicité mathématique lui permettant d'être facilement et efficacement assimilable aux phénomènes mnésiques (Maier & Moore, 2005). En effet, les composantes non-abstraites peuvent

représenter n'importe quel phénomène quel que soit son niveau d'abstraction (ce qui, à cause de son penchant représentationnaliste, n'est pas encore totalement exploité). De plus, en faisant du souvenir une corrélation entre la situation présente et les traces épisodiques en mémoire, MINERVA2 est actuellement toujours le modèle le plus à même de simuler la dynamique de l'appariement global.

Cependant, et malgré sa dynamique globale, MINERVA2 ne prend pas en compte l'évolution dynamique de ses traces en tant que souvenirs. En effet, l'écphorie synergétique stipule que l'utilisation indifférenciée de toutes ses traces est nécessaire, mais l'ordre de présentation des traces n'a ici aucun effet sur l'utilisation de la mémoire : une situation vécue récemment influence autant le système qu'une ancienne, et la temporalité de la dynamique n'est alors pas prise en compte. Il faudrait alors mettre fin à l'aspect statique des traces et à leur indépendance pour que les traces puissent être influencées les unes par les autres et rendre compte ainsi d'une inter-coercition temporelle stable. Nous n'aurions par ailleurs plus besoin de provoquer activement l'oubli des informations mnésiques. En effet, l'oubli ne correspond pas à une perte d'information, mais à une impossibilité de voir émerger un souvenir due à une inadéquation contextuelle (Tulving, 1974). Si le contexte s'y prête, il doit être possible de retrouver une information perdue. L'oubli doit émerger de l'activité dynamique des traces mnésiques en situation.

Le second problème que rencontre MINERVA2 concerne le contenu de ses souvenirs, car celles-ci sont issues uniquement du contenu de la sonde. Or, la mémoire humaine ne retient pas des informations à proprement parler, mais plus exactement la manière dont ces informations sont traitées (Medin & Ortony, 1989). Dès lors, un modèle d'appariement global ne peut pas se contenter d'apprendre directement ce qui a été perçu : il serait plus juste de retenir la manière dont ce qui a été perçu covarie avec le traitement de ces informations (Varela, 1993). Mais pour que ce modèle puisse respecter les qualités énéactives du souvenir, il va falloir se débarrasser de la notion de contenu mnésique au profit d'une réunion processus-contenu : le processus mnésique doit se baser sur des processus mnésiques.

Les deux critiques présentées ci-dessus se rejoignent dans l'idée de rendre les traces sensibles à l'évolution historique du modèle. En premier lieu, la création d'une trace doit être sensible aux états antérieurs du modèle. Percevoir un chien en particulier pour la première fois peut rappeler une morsure dans l'histoire du sujet, et rendre ainsi ce chien menaçant dans le souvenir d'un sujet même sans qu'il n'ait montré d'agressivité. Chaque souvenir est profondément associé à l'histoire du sujet. Pour ce faire, il est nécessaire d'apprendre la covariance sensorimotrice plutôt qu'une représentation de la situation. La production sensorimotrice de la mémoire (l'écho) doit donc fusionner avec la perception sensorimotrice (la sonde) en fonction de leurs corrélations dans la création même d'un nouveau souvenir. Ainsi la mémoire n'apprendra plus des représentations de situations

rencontrées, mais les covariances entre son activité et son environnement.

Deuxièmement, quant à l'utilisation d'une trace, il est nécessaire de la rendre contextuelle. En effet, un souvenir n'influencera pas de la même manière la situation actuelle selon l'histoire d'un sujet. Chaque souvenir particulier n'est pas immuable, il s'adapte et se reconstruit en fonction des situations rencontrées ultérieurement, de la même manière que la mémoire globale s'adapte et se reconstruit en fonction des situations rencontrées (ce qui est le but premier de MINERVA2). Au cœur de la théorie de l'appariement global, le souvenir se comporte donc de la même manière que la mémoire. L'intensité de l'activation d'une trace dépend déjà de sa similarité avec la situation présente dans MINERVA2, mais elle devrait différer selon les situations auxquelles ce souvenir a été confronté (et qui l'on modifié). Les traces mnésiques ne doivent plus être un contenu mais un processus de contextualisation des covariances apprises, une inférence de son activité actuelle en fonction de ses activités passées, ainsi qu'en fonction de l'activité globale de la mémoire.

Enfin, puisque la dynamique d'une trace doit se comporter de la même manière que la mémoire dans sa globalité, il est nécessaire — dans une approche énonciviste radicale — de construire les traces mnésiques à partir d'instances du modèle à une échelle inférieure. Cette particularité donnera à la contextualisation mnésique des caractéristiques fractales.

Afin de rendre ces modifications utilisables, nous allons employer une heuristique probabiliste et mettre au point une formalisation mathématique des changements nécessaires à apporter à MINERVA2 : l'activité d'une trace doit s'inférer à partir de l'activité de l'ensemble des traces, la production mnésique d'une trace doit être contextualisée en fonction de l'historique des interprétations apportées par cette trace, la mémoire ne doit apprendre que des covariances entre sonde et écho, et l'ensemble de ces phénomènes doit faire preuve d'une invariance d'échelle propre à l'utilisation fractale de l'appariement global. Nous avons présenté cette formalisation mathématique et le nouveau modèle de mémoire qui en découle dans l'article suivant intitulé « Modeling an enactivist multiple-trace memory. ATHENA : a fractal model of human memory. » (Briglia, Servajean, Michalland, Brunel & Brouillet, 2017).

2.3 ATHENA : a fractal model of human memory

Modeling an enactivist multiple-trace memory. ATHENA: a fractal model of human memory.

Briglia¹, J., Servajean¹, P., Michalland^{1,2}, A-H., Brunel¹, L., Brouillet¹, D.

¹*Epsilon Laboratory, Paul Valéry University, Montpellier, France.*

²*CNRS-UM LIRMM UMR5506, Montpellier, France*

Abstract

Global-matching models of memory argue that knowledge emerges from the interaction between presented cues and traces of past experiences. But these models generally rely on the use of independent episodic traces, unable to account for global interactions between learned situations (see Versace et al., 2009). Enactivism (Varela, 1993) could theoretically take advantage of an inter-dependant processing of traces to account for abstraction processing using only sensorimotor covariances (Hutto & Myin, 2012), but no mathematical formalization of an enactivist memory has yet been proposed. In this paper, we propose the ATHENA model as an enactivist mathematical formalization of Act-In theories (Versace et al., 2014) within MINERVA2 (Hintzman, 1986) non-specific traces: ATHENA is a fractal model which keeps track of former processes that led to the emergence of knowledge, and is therefore able to process contextual processes (abstraction manipulation). We present three simulations designed to test ATHENA's ability to construct, learn, and manipulate emergent abstractions.

Keywords: Global-matching, fractal, enaction, multiple-trace, memory model

1. Introduction

Since the cognitive revolution in the 1950s, memory has been described as a multistore space (Atkinson & Shiffrin, 1968), that is to say a structural model. One of the main characteristics of structural models of memory is that knowledge is associated with specific memory systems on the basis of different representational forms. Knowledge associated with our motor abilities is thought to be stored in a procedural memory, knowledge associated with our perceptions is stored in the perceptual representation system, episodic memory is thought to contain information associated with a precise spatio-temporal context and semantic memory is thought to contain a fully decontextualized abstract general knowledge (e.g., Tulving's SPI model, 1995).

This multisystem approach of memory has been subjected to various criticisms (e.g., Roediger, Rajaram, & Srinivas, 1990), but among these, the main criticism was based on the independence of the systems. Indeed, the consequence of using independent systems was that the structural models of memory have difficulties accounting for interactions and integrations associated with these various forms of knowledge. This limitation has been at the origin of an alternative con-

cept of memory that appeared in the 1980s, in which representations are not based on their content but on the cognitive architecture: the multiple trace models (Hintzman, 1986; Logan, 1988; Whittlesea, 1987). In these models, a single memory system stores traces of individual experiences. Consequently, these traces are: a) episodic since they reflect the characteristics of the processing episodes that gave birth to them, b) multidimensional since they reflect the different characteristics of episodes. For multiple trace models, knowledge is described as an emergence resulting from the matching between the multicomponent of the present experience and the multicomponent of the past experiences. Therefore, the nature of the knowledge (i.e., episodic or semantic) depends on the result of this matching, which leads to a form of knowledge that is more or less specific to an earlier experience.

However, the multiple trace models have also given rise to a number of criticisms (see Versace et al., 2009, 2014). The main criticism is related to the relations between the traces, more precisely the absence of relations and the consequences this entails. Multiple trace models hypothesize that memory stores traces specific to each experience independently to the other traces of experi-

ences. This independence is both spatial (locatable) and time related (the content of the trace specific to an experience is independent to the content of old stored traces and doesn't influence the content of future traces). But more than anything, this independence of traces does not allow the integration of the traces' contents. That is to say that in multiple trace models, a trace corresponds to a simple co-occurrence of the different components of experiences. However, since Bousfield (1953), or Bransford and Franks (1971), it has been well established that one of the characteristics of human memory is to regroup and integrate information into a coherent whole.

To answer this criticism, Versace et al. (2014) proposed a theoretical model, Act-In, whose main characteristic is to hypothesize that knowledge emerges from the dynamics of two mechanisms: inter-trace activation and multi-component integration. Inter-trace activation involves intra and inter component activation which operate in parallel on the basis of the present experience. Integration is not a simple sum of activations, it is a dynamic non-linear mechanism that operates in parallel with inter-trace activation and progressively constructs knowledge. In opposition of classical multiple-trace models, knowledge is in a state of constant reorganization through the subject activity associated with environmental constraints. But a mathematical formalization of such inter-dependent evolving traces has not yet been proposed, which would be a necessity in order to put Act-In theories to the test with experimental simulations.

To summarize, if multi-trace models challenged the multi-system organization of memory and emphasized an emerging knowledge of the cognitive architecture, their underlying conception of cognition was still inspired by the computer model (i.e., computo-symbolic approach) rather than the brain model (Only Whittlesea, 1989, has proposed a model within a connectionist architecture known as VISA, Variable Integration and Selective Attention). In opposition, the underlying conception of Act-In is similar to the enactive approach of cognition (see Varela, Thomson & Rosch, 1993) which claims that every organism enacts the world in which it exists through its actions. Indeed, Versace et al. (2009) wrote: "...the perception of our environment, just like the knowledge associated with this environment, comes from our activity" (p. 540).

According to enaction, we are not reducible to an isolated brain beside an independent world, we are dynamic entities in constant connection with our environment. The environment is part of the cognitive system. Therefore, cognition does not have to deal with repre-

sentations of the external world, like computo-symbolic and connexionist approaches, but with the direct link to the environment via the sensorimotor system (e.g., O'Regan & Noë, 2001, argue that perceptual experience is built by the mastery of sensorimotor regularities or contingencies). A memory model based on the enactivist theory would then be able to dynamically construct its own structure according to the environment.

According to synergistic ephory (Tulving, 1976, 1982), remembering is a construction at the time of retrieval, so that knowledge emerges from the matching between cues present in the retrieval situation and traces of past experiences. This is essential to formalize an enactivist memory, fundamentally based on the lived experience of its environment (Varela, 1996). To apply synergistic ephory within the enactivist paradigm, we will base our model on the mathematical formalization of synergistic ephory proposed by Hintzman: MINERVA2. Even if this model goes back several decades, MINERVA2 is still widely used due to its ability to simulate a wide range of mnemonic phenomena (see Arndt, 2010) from a mere mathematical formalization of the synergistic ephory principle (Rousset, 2000). Moreover, MINERVA2 uses the same components for its input and output, allowing it to account for motor perception (and more globally for the perception of any cognitive activity). This is a necessity to construct an enactivist memory. However, some modifications are necessary for MINERVA2 to become an enactivist memory model.

First, we must adapt its trace content to an enactivist conception of cognition and use its interactions with the environment through the sensorimotor system (Maturana & Varela, 1987). Covariances between an individual and his/her environment are sufficient to account for cognitive information (Hutto & Myin, 2012). Indeed, if two processes are covariant, they are probably linked. There can be some variability in this probable link, but we can infer it using the covariance intensity *in situ*: this is a contextualization process. Covariance is closely related to an inference function, it can extract a correlation informative to the common origin (or the common purpose) of several processes. In accordance with synergistic ephory, correlation inference between the present situation and memory traces can account for memory emergence. Henceforth, to take into account the inferential peculiarities of covariance, the mathematical formalization of an enactivist memory will be proposed in terms of probabilistic equations.

Second, we must adapt its traces' functioning to the inter-dependent evolving traces of Act-In. Each trace

evaluation must depend on the evaluation of the other traces, and each trace must evolve according to its own activity. A major characteristic of Act-In being a scale invariant integration processing of information (local and global, Versace and al., 2014) is that each trace will have to evolve like memory does in its globality (i.e., learning from the covariances between each situation encountered and its local processing). Given that memory traces have the same functioning as memory in its globality, fractal-type recursive operating should be exploited: some instances of such a memory model should be constructed inside the global model. This characteristic will allow an evolution in the way information is processed by traces, and *a fortiori* by memory.

To our knowledge, an enactivist memory learning sensorimotor covariances has not yet been modeled; we propose in this work such a formalization: the ATHENA model. The theories underlying enaction (sensorimotor cognition without representation) are usually criticized for their lack of applicable formalization and their inability to account for abstraction. ATHENA provides a practical and mathematical implementation of the enactivist theory.

Does memory unconditionally need representations in order to achieve high level processes, or would the complexity of trace dynamics be sufficient for functional equivalents of representations to emerge? Instead of learning retrievable content, an enactivist memory should only learn re-usable sensorimotor processes. This paper aims to analyze how dynamical traces can self-organize in a single-system memory model. In order to achieve this, a probability modeling is used to show how trace dynamics involve a contextual use of mnemonic processing. By modifying MINERVA2 with the trace dynamics proposed by the Act-In model, we will observe a structural similarity between process and content. Memory only learns and processes the memory processes themselves: thus, the ATHENA model acquires a recursive *modus operandi* with a fractal-type scale invariance of memory processes.

To understand what ATHENA offers in comparison to static traces memory modeling, we begin by introducing the MINERVA2 equations with probabilistic equations so that we can develop ATHENA in usable terms for experimental simulations. We conclude with simulations of behavioral data allowing us to put to the test an enactivist memory in abstraction situations for which enactivism has yet to prove its worth.

2. Probabilistic modeling

2.1. MINERVA2

MINERVA2 is a mathematical model of Tulving's (1976, 1982, 1983) synergistic ecphory principle. Ecphory is a synergistic principle in the sense that it matches the cue/stimulus that triggers the search in memory and the content of the totality of traces in memory that are associated with it. It is from this synergy that the awareness leading to a memory is created. In MINERVA2, the cue is characterized by a probe, the traces by a matrix and the recollection by the echo. From the modeling point of view, the echo, denoted by C , is an evaluation of the probe, denoted by θ , on the basis of traces in memory, denoted by X . To explain MINERVA2 in a way that could lead to the probabilistic inference of covariances, we will explain its functioning through probabilistic equations. In this way we will more easily point out what needs to be changed in this model to account for an enactive memory. To do so, index i is used to quantify the different x_i traces for X and index j to quantify the different features C_j and θ_j for C and θ , respectively.

The interpretation of each echo's feature is therefore calculated as $C_j = P(\theta_j)$, where the probability function represents the activity of the memory. However, in order to follow the principles of global matching (i.e. the probe is compared to the totality of traces) and interpret θ_j on the basis of memory functions, MINERVA2 applies the law of total probability. This allows the model to reveal the interpretation of a probe feature in terms of conditional probabilities, such as:

$$P(\theta_j) = \sum_{i=1}^n \left[P(\theta_j|x_i) \cdot P(x_i) \right] \quad (1)$$

where n is the total number of traces. In this manner, the interpretation of θ_j is carried out by adding each interpretation of θ_j according to x_i , weighted by the assessment of the role of x_i in this interpretation. In this model, $P(\theta_j|x_i)$ is the value θ_j learned during the creation of x_i (i.e. x_{ij}) and $P(x_i)$ is the activation value of trace x_i in the situation (i.e. the activation of the trace according to the probe, denoted by A_i).

For MINERVA2, the value of the activation of trace $P(x_i)$ is estimated by comparing probe θ with trace x_i , so that:

$$\begin{aligned}
A_i &= \widehat{P}(x_i) \\
&= \left[\frac{1}{m_r} \sum_{j=1}^m (x_{ij} \cdot \theta_j) \right]^3 \quad (2)
\end{aligned}$$

where x_{ij} is the value of the j feature of trace i , θ_j is the value of the j feature of the probe, m is the total number of features, and m_r is the number of relevant features for the situation (for which the two values, θ_j and x_{ij} , are not zero). $P(x_i)$ is then estimated by observing the similarity between the probe and the trace. To do so, Hintzman has chosen the average multiplication of the two vectors features amplified by cubing. Although relevant in estimating the similarity of integer values, we shall see that the use of floating values allows us to improve the estimation of the activation of a trace, and to use the infimum as a more reliable estimator of likelihood. Here the set of features is tested, and consequently the echo of a feature j depends on the activity of all other features for all of the traces. A component is defined according to all component features, guaranteeing the multi-component aspect of synergistic ephory.

In the end, the calculation of the C_j features of the echo for equation (1) may be performed numerically from the values x_{ij} and θ_j , substituting equation (2) into it so that:

$$\begin{aligned}
C_j &= \widehat{P}(\theta_j) \\
&= \sum_{i=1}^n \left(x_{ij} \cdot \left[\frac{1}{m_r} \sum_{j=1}^m (x_{ij} \cdot \theta_j) \right]^3 \right) \quad (3)
\end{aligned}$$

Echo values are used to account for mnesic construction. While useful for simulations, these values are not part of the learning performed by this model. Finally, MINERVA2 learns a new mnesic situation by adding a new vector to its pool of traces. This vector is a copy of the probe vector θ : a representation of the current situation. The use of the $P(x_i)$ values by Hintzman also makes it possible to determine the intensity $I = P(X)$ of the activity of the model—in other words, the evaluation of the activation of memory X —by adding $P(x_i)$ values as follows:

$$P(X) = \sum_{i=1}^n P(x_i) \quad (4)$$

The calculation of intensity I in equation (4) can also be estimated with equation (2):

$$\begin{aligned}
I &= \widehat{P}(X) \\
&= \sum_{i=1}^n \left[\frac{1}{m_r} \sum_{j=1}^m (x_{ij} \cdot \theta_j) \right]^3 \quad (5)
\end{aligned}$$

The set of features j for each trace i is used to calculate the intensity I . Although convenient to account for the multi-component aspect of memory, this estimate of the intensity is problematic because the evaluation of the activation of a trace depends only on its own activity, without taking into account the different activations of the other traces. In accordance with Act-In theories, we believe that the contextual activation of a trace according to the activity of the memory as a whole would be more efficient: in this manner, the activation of a trace would be relative to the activations of the other traces. Meanwhile, we would also overcome a limitation of MINERVA2 and be able to normalize the $P(X)$ estimate so that its value no longer depends on the number of traces. In order to formalize an enactivist memory within MINERVA2's synergistic ephory, we will need to change its equations to account for sensorimotor covariances and contextualization by inference.

2.2. ATHENA

MINERVA2 inherits its name from the Roman goddess of knowledge. ATHENA reuses the theories that are the foundation of MINERVA2 while applying them in an enactivist way: thus it has been named after the corresponding Greek name of its predecessor.

The first necessary modification must correct the fact that MINERVA2 does not account for contextual dynamics of traces, which is shown by the use of $P(x_i)$ values to weigh $P(\theta_j|x_i)$ values in the echo construction. Though it would be possible, by taking into account inter-trace and inter-component activation (see Versace et al., 2014), to normalize $P(x_i)$ values based on the global activity of the memory—i.e. $P(x_i|X)$. By modulating the calculation of C_j in equation (1) based on the global activity of the memory—while still applying the law of total probability—we obtain the following:

$$\sum_{i=1}^n \left[P(\theta_j|x_i) \cdot P(x_i|X) \right] = P(\theta_j|X)$$

Therefore, instead of considering that the echo is an objective interpretation of the probe (such as $C_j = P(\theta_j)$), the ATHENA echo contextually depends on the

manner in which the probe is processed by the memory: $C_j = P(\theta_j|X)$. Therefore, the calculation of C_j directly affected by X reveals new memory characteristics for the echo:

$$P(\theta_j|X) = \sum_{i=1}^n \left[P(\theta_j|x_i) \cdot P(x_i|X) \right] \quad (6)$$

Equation (6) reveals a similarity of the processes used for the evaluation of a feature by the memory $P(\theta_j|X)$ and the evaluation of a feature by a trace $P(\theta_j|x_i)$. The probability function therefore no longer represents the memory activity exclusively but also the memory content. This similarity between the process and its content leads us to no longer consider the x_i trace as static, but as having the same functioning as the global memory X (so that x_i contains anterior processes traces denoted by x_{ik}). In other words the x_i trace is evolving: its abilities to interpret a situation will depend on the manner in which it has interpreted the preceding situations, just as the global memory can evolve throughout the situations encountered. We can therefore consider that the calculation of $P(\theta_j|x_i)$ is to be carried out with the same formula as $P(\theta_j|X)$ in equation (6), but on a reduced scale:

$$P(\theta_j|x_i) = \sum_{k=1}^o \left[P(\theta_j|x_{ik}) \cdot P(x_{ik}|x_i) \right] \quad (7)$$

where o is the total number of traces x_{ik} inside x_i . This scale invariance of process and content gives ATHENA a fractal-type recursive operation where information is not a objectively retrievable content but a contextual evaluation.

However, to come back to the calculation of C_j , the evaluation of the activity of traces $P(x_i|X)$ must take into account the fact that x_i is included in X . Therefore:

$$P(x_i|X) = \frac{P(x_i \cap X)}{P(X)} = \frac{P(x_i)}{P(X)} \quad (8)$$

Thus, $P(x_i|X)$ represents the evaluation of the trace x_i activity normalized by the evaluation of the memory X activity. The activity of trace x_i is thereby brought to the scale of the global memory activity: small values will be amplified if the activity of memory X is comparatively weak, or will be attenuated if the memory activity is greater. Contrary to what MINERVA2 does, cubing the estimate of $P(x_i)$ to increase the bounds is therefore no longer necessary.

The use of $P(X)$ to normalize $P(x_i)$ must take into account the fact that MINERVA2 and ATHENA are

built on voluntary competition between activations and inhibitions, which translates into positive or negative probabilities that may cancel each other out (see Feynman, 1987, for a detailed explanation of the use of negative probabilities in a dynamic system). The evaluation of the probe is relative and depends on the manner in which the information has been processed. Using only enactivist covariances, we do not have direct access to $P(x_i)$ or $P(X)$ without evaluating them in situation: the model must then infer the evaluation of its own activity.

Firstly, in order to normalize the activity of x_i as a function of the global activity of X , we have chosen to estimate $P(X)$ with a sum of the absolute values of the $P(x_i)$ activations, which makes it possible to take into account the negative activations. Indeed, negative activity must be part of the evaluation of global activity in the same way as positive activity, without canceling each other out:

$$\widehat{P(X)} = \sum_{i=1}^n |P(x_i)| \quad (9)$$

The inference made on the global activity of the memory is representative of an attribution process: the memory subjectively estimates the quality of its interpretation based on the fluidity of the processes carried out (see Whittlesea's SCAPE model, 1993, 1997). Consequently, the calculation of the relative activation of a trace as per equation (8) is made using equation (9) so that:

$$\widehat{P(x_i|X)} = \frac{P(x_i)}{\sum_{i=1}^n |P(x_i)|} \quad (10)$$

Secondly, to account for the central use of $P(x_i)$ in the calculation of $\widehat{P(x_i|X)}$, we use, as does Hintzman, an estimation of the trace/probe interaction able to account for the multi-component aspect of the memory. However, instead of using an estimator of *similarity* between the trace and the probe, the fact that we conceive the trace as a contextual process (and no longer a retrievable content) requires us to choose an estimator of *likelihood* for the probe according to the trace. As with MINERVA2, this estimator must take into account the multi-component aspect of the trace activation, and $P(x_i)$ is estimated using the average likelihood of the features given x_i :

$$\widehat{P(x_i)} = \frac{1}{m} \cdot \sum_{j=1}^m \left[\mathcal{L}(\theta_j|x_i) \right] \quad (11)$$

However, in contrast with MINERVA2, it is no longer necessary to confine ourselves to the use of integers, and

we can use floating values $\in [-1, 1]$ in the model. The role played by a feature is then directly proportional to the relevance of this feature in the situation and in the trace, and no longer requires to define which features are relevant. We can use m (the total number of features) instead of m_r (the number of relevant features used by MINERVA2 in equation (5)).

The ATHENA final equation expresses the interpretation of a feature in situation and is derived from equation (7), so that:

$$\widehat{P(\theta_j|X)} = \sum_{i=1}^n \left[P(\theta_j|x_i) \cdot \widehat{P(x_i|X)} \right]$$

can be developed according to equation (10):

$$\widehat{P(\theta_j|X)} = \sum_{i=1}^n \left[P(\theta_j|x_i) \cdot \frac{\widehat{P(x_i)}}{\sum_{i=1}^n |\widehat{P(x_i)}|} \right]$$

then according to equation (11), to obtain:

$$\begin{aligned} C_j &= \widehat{P(\theta_j|X)} \\ &= \sum_{i=1}^n \left[P(\theta_j|x_i) \cdot \frac{\frac{1}{m} \cdot \sum_{j=1}^m \mathcal{L}(\theta_j|x_i)}{\sum_{i=1}^n \left| \frac{1}{m} \cdot \sum_{j=1}^m \mathcal{L}(\theta_j|x_i) \right|} \right] \end{aligned} \quad (12)$$

This equation is calculated based on the $P(\theta_j|x_i)$ and $\mathcal{L}(\theta_j|x_i)$ values, which need to be inferred using chosen estimators.

$P(\theta_j|x_i)$ is calculated in the same way as $P(\theta_j|X)$ —i.e. according to the estimated effectiveness of each of the traces of x_i included at a lower level. We have seen that, in order to calculate $P(\theta_j|X)$, it is necessary to use an estimator $\widehat{P(\theta_j|X)}$. This particularity is again found on a lower scale, by estimating $P(\theta_j|x_i)$ using equation (12) applied to the calculation of $\widehat{P(\theta_j|x_i)}$.

$\mathcal{L}(\theta_j|x_i)$ represents the likelihood of a probe feature according to a memory trace: it is an evaluation of familiarity that determines to what extent x_i makes it possible to effectively interpret θ_j . As with $P(\theta_j|x_i)$, the likelihood $\mathcal{L}(\theta_j|x_i)$ is calculated based on the estimated effectiveness of each trace included in x_i at a lower level. Therefore, $\widehat{\mathcal{L}(\theta_j|x_i)}$ must take into account the total intensity of $P(x_{i_k})$ values but cannot employ their sum, due to the risk of encountering the same problem as with MINERVA2 while calculating $P(X)$. Indeed, the evaluation of the activation of a trace should not depend only on its own activity, without taking into account the different activations of the other traces: Traces shouldn't be independant. It is therefore necessary to

normalize $P(x_{i_k})$ values according to $P(x_{i_k}|x_i)$ values by adapting them to the particularities of negative probabilities, so that:

$$\widehat{\mathcal{L}(\theta_j|x_i)} = \sum_{k=1}^o \left[|P(x_{i_k})| \cdot P(x_{i_k}|x_i) \right] \quad (13)$$

Equation (13) makes it possible to estimate the likelihood of a feature processed by a trace according to past activities of this trace contextualized by the processing of the current situation. It is calculated by applying equations (10) and (11) from likelihoods of a lower scale, such as $\widehat{\mathcal{L}(\theta_j|x_i)}$ equals to:

$$\sum_{k=1}^o \left[\left| \frac{1}{p} \cdot \sum_{l=1}^p \mathcal{L}(\theta_{jl}|x_{i_k}) \right| \cdot \frac{\frac{1}{p} \cdot \sum_{l=1}^p \mathcal{L}(\theta_{jl}|x_{i_k})}{\sum_{k=1}^o \left| \frac{1}{p} \cdot \sum_{l=1}^p \mathcal{L}(\theta_{jl}|x_{i_k}) \right|} \right] \quad (14)$$

where x_{i_k} is the trace k present inside the trace x_i , θ_{jl} is the feature l in the feature θ_j (corresponding to an ATHENA instance), and p the total number of features θ_{jl} inside θ_j . The decomposition of θ_j could seem counter intuitive, but ATHENA being a fractal model, the ATHENA instances embedded inside the traces of the global ATHENA gather some features according to their correlations. According to enactivism (varela, 1993), covariances permits to gather several processes into one, so meaning can be constructed by correlation (this is a fundamental basis of constructivism, Piaget, 1936, 1937). Each trace has a different number of features according to fractal gathering, and these gatherings need to be decomposed in order to calculate $\widehat{\mathcal{L}(\theta_j|x_i)}$. A practical explanation of this kind of complex constructions will be provided in chapter 3.

Henceforth, in order to facilitate understanding of the indices we can use the scale invariance to define equation (13) at a higher scale:

$$\widehat{\mathcal{L}(\theta|X)} = \sum_{i=1}^n \left[|P(x_i)| \cdot P(x_i|X) \right] \quad (15)$$

and describe the estimator of equation (14) at a higher scale in terms of a memory X (consisting of x_i traces) and a situation θ (consisting of θ_j features), making it possible to define the calculation of intensity I for ATHENA:

$$\begin{aligned}
I &= \widehat{\mathcal{L}(\theta|X)} \\
&= \sum_{i=1}^n \left[\left| \frac{1}{m} \cdot \sum_{j=1}^m \mathcal{L}(\theta_j|x_i) \right| \cdot \frac{\frac{1}{m} \cdot \sum_{j=1}^m \mathcal{L}(\theta_j|x_i)}{\sum_i \left| \frac{1}{m} \cdot \sum_{j=1}^m \mathcal{L}(\theta_j|x_i) \right|} \right] \quad (16)
\end{aligned}$$

Finally, the multi-component likelihood of the probe given the memory (i.e. the familiarity of the situation) in equation (16) is calculated based on the multi-component likelihoods of equation (11), so that $\widehat{\mathcal{L}(\theta_j|x_i)}$ has the same fractal qualities as $\widehat{P(\theta_j|x_i)}$.

The ATHENA model contains ATHENA instances processing information in the same manner as the global model. However, given that ATHENA is a constructivist model (i.e. it evolves according to interactions with its environment), this type of recursion cannot be infinite. We shall subsequently see how learning is built, but first we must generalize these estimators for a concrete application of values x_{ij} and θ_j .

As with MINERVA2, the estimator of $P(\theta_j|x_i)$ represents the value of the feature j learned by trace x_i (i.e. $\widehat{P(\theta_j|x_i)} = x_{ij}$) with the difference that x_{ij} is a representational content for MINERVA2 and a contextual process for ATHENA (it makes sense only while used with all other x_{ij} values). With respect to the estimator of $\mathcal{L}(\theta_j|x_i)$, the familiarity principle is expressed by an optimization equation which aims to minimize x_{ij} over θ_j (i.e. minimizing free energy according to Friston's theories—Friston, 2010; Friston, Kilner & Harrison, 2006). To achieve this, we calculate the optimal value of the target function x_{ij} over the admissible value θ_j (i.e. the calculation of the infimum that we have adapted to the logic of negative probabilities):

$$\begin{aligned}
\widehat{\mathcal{L}(\theta_j|x_i)} &= \inf x_{ij}(\theta_j) \\
&= \frac{|\theta_j + x_{ij}|}{2} - \frac{|\theta_j - x_{ij}|}{2} \quad (17)
\end{aligned}$$

The scale invariance of $P(\theta_j|X)$ shows us that the content of x_i traces is an activity. Therefore, the learning achieved by each trace does not merely memorize information as content but rather as a process. While MINERVA2 memorizes only the information included in the probe, ATHENA memorizes the manner in which the information has been processed. In this way, during the creation of a new trace, MINERVA2 reuses the raw θ_j values (as the new x_{ij} value for the new trace i) without taking into account the C_j values of the echo. However, in order to memorize the activity of the model

and therefore reveal the enactivist character of memory, it is necessary to compare θ_j and C_j ($\forall j$).

If θ_j and C_j are similar, the learned value must reveal the reinforcement carried out by memory on the evaluation of the situation. If they are disparate, the learned value must reveal the existing inadequacy between the evaluation and the situation. If one of the two values is zero, the learned value must reveal either the new perceived information (if C_j is zero) or the interpretation carried out (if θ_j is zero). Finally, if the two variables are zero, the learned value must also be zero.

Consequently, we calculate the new trace feature value x_{ij} such as:

$$x_{ij} = \frac{\theta_j + C_j}{2} \quad (18)$$

While learning the average value between θ_j and C_j , ATHENA learns the effectiveness of how the information has been processed. In this sense, the memory content is not a datum per se, but becomes meaningful only when compared to the situation and to the totality of memory traces. Therefore ATHENA becomes an enactivist model, able to adapt itself to the peculiarities of its interactions with the environment.

The fractal integration of new situations based on the way in which they are processed reveals two essential characteristics in the operation of ATHENA: (i) the new traces are created based on the old ones (as the old traces are the ones that makes it possible to interpret the new ones and this interpretation is a part of memory); and (ii) the old traces are modified based on new situations (due to the fractal modus operandi of the model, the preceding traces learn how they have made it possible to learn the new ones). In the end, each memorized situation is distributed all over the existing traces, and can account for Act-In inter-traces and inter-component activation and integration. Due to the fractal matching between the ongoing situation and the previously learned ones, an emerging memory carries all the data learned in connection with its current situation. In contrast with MINERVA2, ATHENA traces are not independent, and more complex processes between one trace and the others may emerge from it. We shall verify this using experimental simulations.

We have seen that the estimation of the likelihood of the situation based on the memory $\widehat{\mathcal{L}(\theta|X)}$ (or the *Intensity* of the activity of the model that replaces $P(X)$ in MINERVA2) is carried out based on x_i and θ_j . Likewise the estimation of the likelihood of a feature according to a trace $\widehat{\mathcal{L}(\theta_j|x_i)}$ is carried out according to x_{ik}

and θ_j . This type of estimator makes it possible to evaluate the familiarity of the situation based on the memory activity and summarizes the operation of ATHENA. It is a more reliable indicator than $P(X)$ to synthesize memory activity, and it also makes it possible to account for a characteristic of human memory identified since Bartlett (1933): all memories consist of a construction and a reconstruction. Indeed, the fact that a value of θ_j is expressed with values of θ_{j_i} shows that the evaluation of the features in the current situation is carried out in a specific manner according to the traces: a trace may combine θ_{j_i} features to use them as a single θ_j feature.

To explain this, we can take this expression to a higher level and see how $\widehat{\mathcal{L}(\theta|X)}$ summarizes the activity of the model. $\widehat{\mathcal{L}(\theta|X)}$ considers θ_j values as a whole, forming the situation in its multi-component complexity and x_i values as a whole forming the memory in its synergistic complexity. The main role of this model is to bring together different features to evaluate the whole of the dynamic interactions of the memory. To combine θ_j values as a whole implies that θ is evaluated as a set of correlations linking the different components of the situation. The use of θ postulates that the different features are correlated and that it is necessary to infer these correlations according to previously learned situations. This principle makes it possible to account for inter-component integration (i.e. between the features of the trace), which MINERVA2 did not, as it could only account for inter-trace integration. Henceforth, ATHENA applies Act-In integration process to MINERVA2 (see Versace et al., 2014).

$\widehat{\mathcal{L}(\theta_j|x_i)}$ functions in the same way and accounts for the existing relationships in a set of features whose correlation was effective at the moment of learning. During the activation of a trace, using an instance of ATHENA or a formula to calculate $\widehat{P(\theta_j|x_i)}$ and $\widehat{\mathcal{L}(\theta_j|x_i)}$ depends on the number of features concerned. If θ_j is made up of a single feature, we shall use x_{ij} and $\inf x_{ij}(\theta_j)$ to calculate, respectively, $\widehat{P(\theta_j|x_i)}$ and $\widehat{\mathcal{L}(\theta_j|x_i)}$. If, however, θ_j is made up of several features, we have an instance of ATHENA capable of evaluating the correlation links between these features, and we proceed with the recursion of the equations to obtain an effective result.

Finally, in an enactivist memory, correlations are the only relevant information: correlations between probe and echo, inter-trace correlations, inter-component correlations, and correlations between echo features. The creation of an ATHENA instance during the creation of

a trace depends on the manner in which the model has evaluated the corresponding situation. By calculating C_j values, ATHENA determines if the interpretations of the features result in correlations between them. The correlated features make it possible to create one (or more) instances of ATHENA in the new trace. This determination of correlations is carried out based on Gestalt criteria of spatial proximity and intensities during the same time event: nearby echo features sharing similar values shall be considered correlated, while spatially disparate values or values with overtly different intensities will indirectly form demarcations between ATHENA instances and/or features interpreted as autonomous.

The threshold at which the values of features are considered similar shall be based on the familiarity of the situation—i.e. by reusing intensity $\widehat{\mathcal{L}(\theta|X)}$. If the processing operation carried out is not fluid, the familiarity is uncertain and $\widehat{\mathcal{L}(\theta|X)}$ will approach zero. Therefore, interpretation is difficult and in order for features to be considered correlated, it is necessary for them to be truly equal. However, if the fluidity of the process is high in intensity, the familiarity judgment will be more certain (positive or negative) and the absolute value of $\widehat{\mathcal{L}(\theta|X)}$ will approach 1. Therefore, the interpretation will be easier and features having broadly close values shall be considered correlated.

In the end, intensity $\widehat{\mathcal{L}(\theta|X)}$ and echo $\widehat{P(\theta_j|X)}$ —equations (16) and (12)—are both estimators based on other estimators. Indeed, ATHENA does not directly access information objectively stored in memory, because the traces contained in memory are processes that do not hold meaning independently of the evaluation of the whole situation. The model uses these processes to estimate their relative activities and interpret the relevance of the obtained memory pattern. This pattern is conceived here and now, but in order to attribute to this activity the character of memory, it is necessary for the mnemonic activity to be judged as *past*. The memory may only access its own activity, and therefore must infer its own qualities based on its contextual activity: it is an attribution process that is here represented by the use of estimators.

2.3. Conclusion

The mathematical formalization of ATHENA meets the requirements of an Act-In implementation of an enactivist memory: memory traces are no longer independent, they are contextualized and transformed as experiences are lived, and only covariances stand for information.

To summarize the operation of ATHENA, we have developed two estimators: the intensity $\widehat{\mathcal{L}(\theta|X)}$ —i.e. a quantification of familiarity—allowing the evaluation of the likelihood of a situation θ based on the memory X ; and the echo $\widehat{P(\theta_j|X)}$ —or the output of the system—allowing the interpretation of a situation θ based on memory X for a given j feature.

The principal characteristic of memory that ATHENA models is that its structure is constantly modified during its use, allowing it to account for a characteristic of memory that has been known since Bartlett (1933): its constructed and reconstructed character. As a consequence, past situations will no longer be the only factor determining the interpretation of the current moment, but also possible interactions at the present time between all forms of interpretations already encountered. In this manner, each new trace is linked to the others as a whole and evolves on the basis of the next ones.

In other words, in ATHENA, each trace learns how the information that led to its creation has been processed—it constructs itself—as well as how it helps to process incoming information—it reconstructs itself. In the same manner, the memory (i.e. the totality of the traces) learns how the encountered pieces of information have been processed—it constructs itself—as well as how it continues to process the information—it reconstructs itself continually.

3. Modeling simplification

The probabilistic explanation developed in the preceding chapter has made it possible to account for the theoretical requirements to model an enactivist memory: the synergistic ephory requirements offered by MINERVA2, the inter-dependent evolving traces required by the implementation of Act-In, and the sensorimotor covariances required to be learned by an enactivist system. But it is now necessary to explain them in a more practical way in order to program such a memory model.

3.1. Synergistic ephory

While modifying MINERVA2 equations, there are several part of its functioning that we kept to keep track of a synergistic ephory memory. In brief, we kept its structure. Memory is a set of vectors (traces), for which each trace is compared to the present situation (the probe, a vector as well). These comparisons give rise to a list of similarity values (one for each trace),

which permits the calculation of a list of activations values (one for each trace as well). Activation values provide an evaluation of the memory activity (the Intensity value), along with an evaluation of the memory construction (the echo defined as a vector). MINERVA2 and ATHENA both fall back on the law of total probability, respectively, with equations (1) and (6). Therefore, memory (i.e. the echo) is constructed from the synergy of the comparison of each trace with the present situation (weighted according to their activations), and is therefore defined as a synergistic ephory process.

3.2. Sensorimotor covariances

To account for an enactivist memory, ATHENA does not learn the mere probe which it was presented to. It learns instead the correlation information between probe and echo: their average values. A new trace will be constructed from a vector of feature covariances, as with equation (18). Furthermore, to account for sensorimotor processings covariances, a new trace will not only be constituted of numeric values but of internal processing of these values as well: one or several new ATHENA model(s) may be constructed inside this trace. To determine whether or not several covariances values shall be reunited, the absolute intensity value is used as a correlation interval (so the correlation strength is inferred from the memory fluency). For example, if the average values between probe and echo gives the vector (0.8, 1, -0.5, 0.2, 0, -0.1) and the intensity value is 0.3, the new trace will be constructed as ([0.8, 1], -0.5, [0.2, 0, -0.1]). This vector contains three object: an ATHENA, a numeric value and another ATHENA (a single component value doesn't give rise to an ATHENA because the purpose of the model is to produce multi-component correlations: only a group of components interpreted as correlated can merge in an ATHENA instance). Each of these ATHENAs constructed inside the global ATHENA starts with the features vector that gave rise to its creation, and will evolve according to its future activities. Therefore, sensorimotor covariances (defined as processes and not mere content) will be able to adapt themselves to new situations.

3.3. Contextualization by inference

Determining how to use such sensorimotor covariances instead of retrievable content was more challenging. Most of the equations developed in this paper were intended to determine it. Since the fractal character of ATHENA requires the use of the same equations for the global model as for the instances of ATHENA inside the traces, we must first determine how activations are

calculated. Each trace activation must be inferred according to the others, so we use similarity values (S_i) in equation (8) such that:

$$A_i = \frac{S_i}{\sum_{i=1}^n |S_i|} \quad (19)$$

where n is the total number of traces. This inference will be able to determine both the intensity value and the echo values. The intensity value is calculated in equation (13) and (16) from a sum of all absolute similarities weighted by their corresponding activations, such as:

$$I = \sum_{i=1}^n (|S_i| \cdot A_i) \quad (20)$$

The echo values are calculated in equation (6) and (12) with a sum of all trace values weighted by their corresponding activation, such that:

$$C_j = \sum_{i=1}^n (T_{ij} \cdot A_i) \quad (21)$$

where T_{ij} is the feature j of trace i . This equation is similar to MINERVA2's echo calculation (as they both stem from the law of total probabilities), but S_i and A_i are quite different from one model to the other. For ATHENA, A_i stands for an inference able to contextualize the echo and intensity calculations. And S_i does not determine a similarity between the probe and a trace, but a likelihood as has been seen in equation (11):

$$S_i = \frac{1}{m} \cdot \sum_{j=1}^m \mathcal{L}_j \quad (22)$$

where m is the number of features in the trace, and \mathcal{L}_j is a likelihood function relevant for j . But m does not have the same value for all traces, and the likelihood equation may change as well, as we will now show with regard to the fractal peculiarities of ATHENA.

3.4. Fractal covariances

To understand more intuitively the fractal characteristics of ATHENA, a simplification of the model is presented in Figure 1. This figure consists of a probe ($\theta_1, \theta_2, \theta_3, \theta_4$), an echo (C_1, C_2, C_3, C_4), a trace (consisting of a numeric value $T_{1,1}$ and an ATHENA instance $T_{1,2}$ inside the trace), a similarity S_1 (for simplification purposes, activation is not presented in this figure), and an intensity I . The ATHENA instance present within the trace has the same specificities (with a probe consisting of the three last values of the probe vector).

To calculate S_1 , we saw that equation (22) uses a likelihood function that can be defined as function $\mathcal{L}(x; y)$ where x is a feature of the trace and y is a feature of the probe. This likelihood function has yet to be defined. The trace in Figure 1 has only two features: a numeric value $T_{1,1}$ and an ATHENA $T_{1,2}$ (which has been constructed according to the correlation between covariances of θ_2, θ_3 and θ_4 contextualization upon this trace construction, as seen previously). In this example, the similarity value must be calculated such as: $S_1 = \frac{1}{2} \cdot (\mathcal{L}(T_{1,1}; \theta_1) + \mathcal{L}(T_{1,2}; (\theta_2, \theta_3, \theta_4)))$. Indeed, the construction of an ATHENA instance refers to features that have been interpreted as correlated, so we need to define how this ATHENA instance can interpret the current situation according to its previous interpretations. $\mathcal{L}(T_{1,2}; (\theta_2, \theta_3, \theta_4))$ refers to the fluency of this ATHENA instance while interpreting $(\theta_2, \theta_3, \theta_4)$, namely its intensity. To calculate S_1 , we need to inject the corresponding probe parts into the ATHENA instance, and get back its intensity value. But then, we need also to calculate $\mathcal{L}(T_{1,1}; \theta_1)$ according to equation (17) such as: $\mathcal{L}(T_{1,1}; \theta_1) = \frac{|\theta_1 + T_{1,1}|}{2} - \frac{|\theta_1 - T_{1,1}|}{2}$. In the end, $S_1 = \frac{1}{2} \cdot ((\frac{|\theta_1 + T_{1,1}|}{2} - \frac{|\theta_1 - T_{1,1}|}{2}) + intensity(T_{1,2}))$ where $intensity(T_{ij})$ is a function referring to the calculation of intensity I of an ATHENA instance inside T_{ij} feature with equations (20).

The calculation of echo values in equation (21) uses the T_{ij} value. But T_{ij} can be either a numeric value (in which case it will be used as it is), or an ATHENA instance (in which case its echo values will be used as T_{ij} values). We just saw that intensity is calculated from ATHENA instances' intensities (recursively until only numeric values are reached), echo is likewise calculated from ATHENA instances' echoes (until numeric values are reached). These fractal recursions of mnemonic activities are not limited to a unique level of ATHENA instances within the global ATHENA model. They indeed occur within ATHENA instances inside other ATHENA instances. But this kind of recursion is not infinite; it deepens progressively as probes are presented. ATHENA starts with a unique level of traces; it will then become more complex through the construction of new ATHENA instances while correlations are interpreted. This occurs in the same way as natural fractal constructions do (e.g. romanesco broccoli growth).

4. Experimental Simulations

To test the efficiency of an enactivist multiple-trace memory in abstraction, this section puts ATHENA in situations for which enactivism has yet to prove its

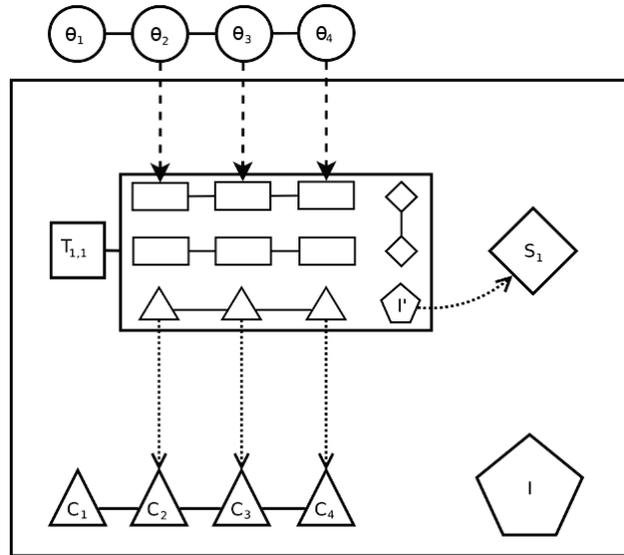


Figure 1: ATHENA fractal pattern simplification

worth. The experimental simulations were set in order to compare ATHENA (an enactivist model) and MINERVA2 (its non-enactivist equivalent). For this part, we shall introduce simulations of three experimental protocols. The first simulation, based on the DRM paradigm (see Deese, 1959; Roediger & McDermott, 1995), was considered in order to ensure that ATHENA maintained the basic abstraction capacities of MINERVA2. The second aimed at simulating a priming effect. To do this, we chose to replicate the results observed in Heurley et al.’s (2013) experiment. In this way, we could directly test ATHENA’s capacity to learn an abstraction process as if it were a retrievable content. Finally, simulation of the Brouillet et al. (2015) experiment allowed us to test ATHENA’s ability to manipulate emerging abstractions as if they were mere retrievable content.

Implementing ATHENA required a computer environment that would intuitively allow for the construction of a fractal global-matching memory process. In consequence, this environment was required to support some dynamical changes and had to be able to use its own processes as objects to construct new states. We chose the Smalltalk language (designed for constructionist learning and human-computer symbiosis) in the Pharo environment. In this way, we were able to use processes per se as data, to ease the fractal use of this model and to modify the source code on the fly according to our experimental needs. The ATHENA open-source code is available on this SmalltalkHub repository:

<http://smalltalkhub.com/#!/~JohanBriglia/>

ATHENA

4.1. Simulation of the DRM Paradigm

Original experiment. The simulation of the DRM paradigm (see Deese, 1959; Roediger & McDermott, 1995) summarizes memory capacities for recognition judgment as per the MINERVA2 model. In this task, participants learned some lists of words (e.g. *bed, rest, awake*, etc.) linked to a critical lure that was not presented (e.g. *sleep*). During a recognition task, the participants had to state whether the given words were presented during the learning phase. Four types of word were presented in the recognition list: words seen previously (i.e. old related), the critical lure, new words weakly related with the critical lure (i.e. new related), and new words unrelated to the critical lure (i.e. new unrelated). The results showed on one hand that the critical lure was as well recognized (sometimes even better) as old words, but also that new related words were more subject to false alarm than new unrelated words, yet less so than the critical lure. This experiment demonstrated that memory recognition is an emerging construction from a global matching process (see Arndt, 2010, for a discussion).

Simulation procedure. To simulate this experiment in a manner adapted to MINERVA2 and ATHENA, we produced the necessary probe categories in accordance with the methods developed by Hintzman (1986, 1988). To do so, each probe θ included 200 features with values

$\in \{-1, 0, 1\}$. To generate the critical lure and new unrelated probes, we took seven random probes with different features: one of the seven was randomly chosen as the critical lure, and the six others made it possible to present new elements without connection to the critical lure. Each old related probe was generated by randomly modifying 5 per cent of the critical lure features for a different random value $\in \{-1, 0, 1\}$; this operation was carried out six times to obtain six old related probes. The generation of new related probes reused the same principle but since these only required weak correlation to the critical lure, the random modifications took place for 70 per cent of the features. The six old related probes were presented during the learning task. The critical lure as well as a randomly chosen old related probe, new related probe and new unrelated probe, were presented during the recognition task. The quantification of familiarity represented by intensity I was analyzed in the two models—i.e. $P(X)$ for MINERVA2 and $\widehat{\mathcal{L}(\theta|X)}$ for ATHENA—so that it may be compared to the old/new recognition performance observed in the classic experiment. Thus, obtaining a positive intensity corresponded to an *old* response whereas a negative one corresponded to a *new* response. The experiment was simulated 5,000 times for the purposes of the statistical analysis. The results obtained in the classic experiment (Roediger & McDermott, 1995), as well as those obtained with both models (MINERVA2 and ATHENA) are described in Figure 2, showing the hit and false alarm rate (i.e. the probability of answering *old*) for all the conditions.

Results analysis. First of all, we can see in Figure 2 that the global pattern of responses is about the same for the classic experiment and the simulations carried out by the two models. In fact, both models are effectively able to judge the probes as *old* or *new*, with a hit rate of 100 per cent for old related and a false alarm rate of 9.24 per cent for new unrelated for MINERVA2 ($\chi^2(1) = 8308.312$, $p < 0.001$), and a hit rate of 100 per cent for old related and a false alarm rate of 13.38 per cent for new unrelated for ATHENA ($\chi^2(1) = 7639.795$, $p < 0.001$). Then, we observed significantly more false alarms (respectively, 23.14 per cent and 25.60 per cent) for new related words compared to new unrelated (for MINERVA2, $\chi^2(1) = 355.981$, $p < 0.001$, and for ATHENA, $\chi^2(1) = 237.914$, $p < 0.001$), which is consistent with a recognition error linked to a global matching process. However, the most spectacular false alarm rate was observed for the critical lure with a 100 per cent rate of false alarms for the two models, which was higher than false alarm rates observed for new related probes ($\chi^2(1) = 6241.676$, $p < 0.001$

with MINERVA2, and $\chi^2(1) = 5923.567$, $p < 0.001$ with ATHENA). The false alarm rate of the critical lure probes was remarkably similar to the hit rate for the old related probes for both models. These results were similar to those obtained in the original experiment and were consistent with the DRM paradigm.

Finally, the response patterns were quite consistent between models $\chi^2(3) = 0.79$, $p = 0.85$. With this first simulation, we can say that the modifications introduced in ATHENA have not altered its ability to produce the recognition pattern observed with MINERVA2. Thus, even with an enactivist modus operandi, ATHENA inherits the abstractions abilities of MINERVA2.

4.2. Simulation of a Perceptual Priming Effect

With the first simulation, we showed that ATHENA kept the essential abstraction characteristics of MINERVA2, so enactivism is indeed able to perform an abstraction emergence. Now we shall examine the advantages resulting from the enactivist character of ATHENA (i.e. the memory of its process history). To do so, we decided to employ a priming effect procedure. Indeed, the study of a perceptual priming effect (see Heurley et al, 2013) allowed us to test the relevance of the principle according to which the memory system codes the manner in which it has processed the information.

Original experiment. In this experiment, the participants had to perform a color discrimination task on a color patch (i.e. yellow or green). For each trial, the color patch was preceded by a visual prime (i.e. the name of an object typically associated with the color green, such as a frog, or with the color yellow, such as a banana). Results showed that in the case of congruency between the prime and the target (e.g. the word *frog* followed by a green patch), the response time was faster than in case of an incongruent trial (e.g. the word *frog* followed by a yellow patch). This experiment showed that the processing operation carried out during the presentation of the word played the role of a prime and had an influence on the color discrimination task. Thus, throughout the simulation of this experiment, we were able to confirm whether the processing operation carried out by the models were part of the learning.

Predictions. Because MINERVA2 is not able to code its own activity and to use it efficiently, there was no reason to predict any difference in terms of level of intensity depending on the trial condition (i.e. congruent or incongruent). However, because ATHENA is able to

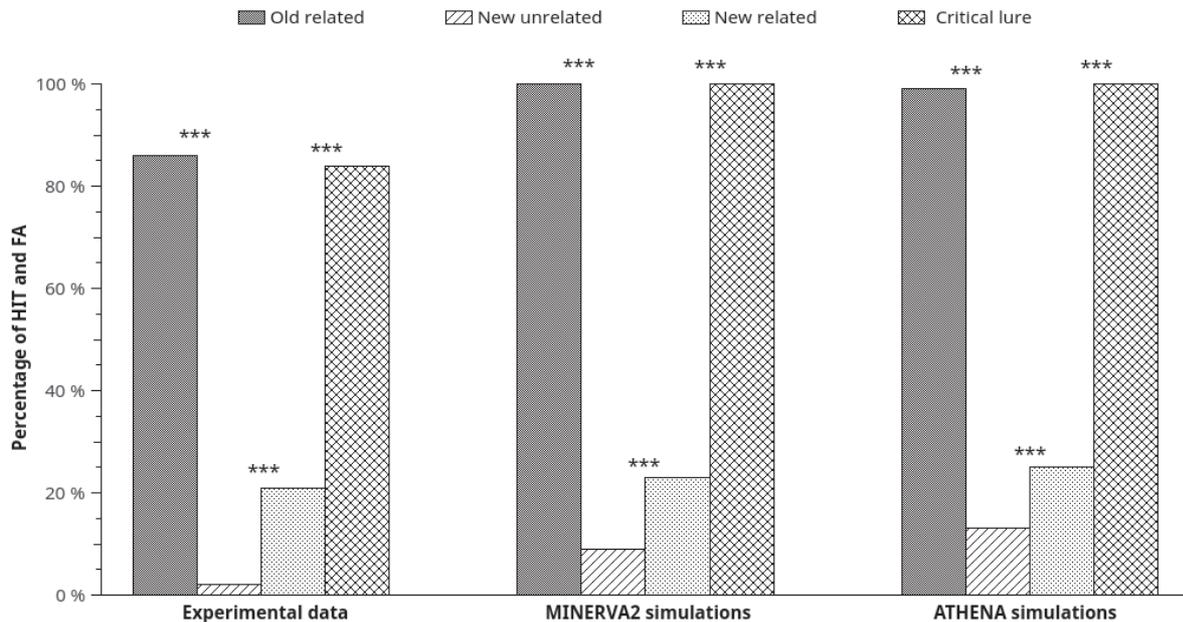


Figure 2: DRM. Percentage of Hit and False Alarm according to the nature of the item from experimental and simulation data

dynamically adapt its activity to both the current situation and former situations, we predicted that the level of intensity would depend on the trial condition (i.e. with a higher intensity for a congruent trial than for an incongruent one).

Simulation procedure. For this simulation, MINERVA2 and ATHENA had first to learn a list of random word-color associations (to ensure the subject’s cross-dimensional perceptual knowledge: word shape and object color); they were then engaged in a priming procedure, with a prime followed by a categorization task of one of the colors learned. For instance, the learning task may include the presentation of the color yellow and the word *banana* synchronously, then the presentation of the color green and the word *frog*, and so on. Then, both models were engaged in a color categorization task (e.g. green or yellow), preceded by a word that is supposed to elicit a color (e.g. *banana* or *frog*). For both models, we compared situations in which the prime and the target were congruent to those where they were incongruent.

In practice, the two categories of word and color were each depicted by a vector of four features. Four different words and four different colors were generated with random values for each feature $\in \{-1, 1\}$. This operation was repeated until each word and each color had different values to the other items in its category. The probes

θ used by the models then had eight features, the first four corresponding to the word and the last four to the color. For instance, if the word *banana* was determined by the features vector (1, 1, -1, -1) and the color yellow by the features vector (1, -1, -1, 1), then the probe associating the word *banana* with yellow was determined by the vector (1, 1, -1, -1, 1, -1, -1, 1).

During the learning phase, the models learned four items: four unique word-color associations presented in random order to both models. For the two conditions of trial (i.e. congruent or incongruent), the prime consisted of the presentation of one of the words and had the color features set to zero—for instance, the vector (1, 1, -1, -1, 0, 0, 0, 0) for the word *banana*. Then, during the categorization task, the presented probe included the features of the color, and the features for the word were set to zero—for instance, the vector (0, 0, 0, 0, 1, -1, -1, 1) for the yellow color.

In the end, the color categorization obtained in echo C was always correct for both models (100 per cent correct responses), as observed in Heurley et al.’s (2013) experiment. However, we were more interested in the ease with which these categorizations took place. To analyze the simulation results we kept the value of $\widehat{P}(X)$ for MINERVA2 and $\widehat{\mathcal{L}}(\theta|\widehat{X})$ for ATHENA (in the two cases these are the I intensity value from the models, from equations (5) and (16), respectively), which made

it possible to effectively estimate the intensity of the categorization: in comparison with the observed data in Heurley's experiment (Figure 3a), higher intensity corresponded to higher familiarity, and thus to a faster response time (i.e. intensity is inversely related to response time). To facilitate the reading of the results, we reversed the average intensity indicator (i.e. $1 - I$), which allowed us to directly compare response patterns (Figure 3b). The simulation was reproduced 5,000 times for the two models due to statistical requirements, and the full results are presented in Figure 3.

Results analysis. First of all, the general pattern of results was similar between the original experiment and the ATHENA simulations, and at the same time quite different from the MINERVA2 simulations. Using t tests for independent samples, we showed that the intensity observed with ATHENA was significantly higher in the congruent situations than in the incongruent ones: $t(9998) = 8.560$, $p < 0.001$. No differences were observed between those conditions for the MINERVA2 simulation: $t(9998) = 0$.

Although word-color associations were learned for both models in the learning phase, during the priming phase, the behaviors for both models were quite different. As predicted, only the ATHENA model was able to reproduce this priming effect pattern. Indeed, even if the prime perception provoked for both models the simulation of the color associated with the word (e.g. the perception of the word *banana* recalls the yellow color in the echo), only the ATHENA model learned this memory activity. In the case of MINERVA2, where only the external stimulus was learned (the prime), this processing operation was not memorized. Therefore, the color categorization was not modified by the prime formerly presented, and no difference in memory processes occurred between congruent and incongruent situations. MINERVA2 could only simulate this experiment if a compound cuing (Ratcliff & McKoon, 1988) between probe and echo was re-injected in the model, but this would be specific to this experiment and would modify this model's results in other simulations. MINERVA2 would need an external intervention to adapt to the situation. However, in the case of ATHENA, the fact that the word (e.g. *banana*) triggered a memory of the color (e.g. yellow) was memorized as a correlation. As for any repetition, the color emergence during the prime presentation facilitated the following categorization task in a congruent situation (or hindered it in an incongruent situation). Yet, this was not the repetition of a retrievable content, but the repetition of a mere process. The priming effect occurred only with ATHENA due to

its ability to learn its memory processes, and hence its memory historicity. ATHENA processing of the situation can be learned as if it was a retrievable content within its standard operating procedure. An enactivist memory can automatically adapt its processing to the situation.

4.3. Simulation of abstractions interaction

With the second simulation, we provided a significant improvement of ATHENA compared to MINERVA2. Indeed, we showed that ATHENA keeps track not only of former experiences but also of the processes themselves, so that we could simulate a priming effect. Enactivism is not only able to perform an emergent abstraction, it can learn these emergent abstractions too. But can ATHENA use these emergent abstractions just like regular models can use their retrievable abstractions? The last simulation is dedicated to highlighting the benefits of using a fractal structure in this memory model. We thought that the experiment designed by Brouillet et al. (2015) could be adapted to fulfill that objective.

Original experiment. In this experiment, participants learned a series of common object names and then performed a recognition task on a series of object images. During the recognition task, the authors manipulated the orientation of the objects' graspable part so that half of the objects was oriented to the left and the other half was oriented to the right. Such manipulation was designed to introduce a form of compatibility between the object to be recognized and the laterality of the action used in order to perform the recognition task (see Tucker & Ellis, 1998). Thus, there were situations in which the nature of the object (i.e. old or new), the orientation of the graspable part (i.e. left or right) and the corresponding position of the response key (i.e. response *old* on the left or on the right) were compatible or incompatible. The authors showed some additive effect of the nature of the item and the stimulus-response compatibility. In other words, participants responded *old* significantly more often for old items than for new ones and they also responded *old* significantly more often in compatible situations (i.e. *old* response-key in the same position as the orientation of the graspable part of the object) than in incompatible situations (see Figure 4). Thus, the effect might be observed because participants had knowledge about how to grasp an object, but also about the link between objects' visuals and their names. During recognition, both effects were observed and had a direct influence on recognition performance.

Thus, answering *old* to an image, where only names were learned, required adaptation between several

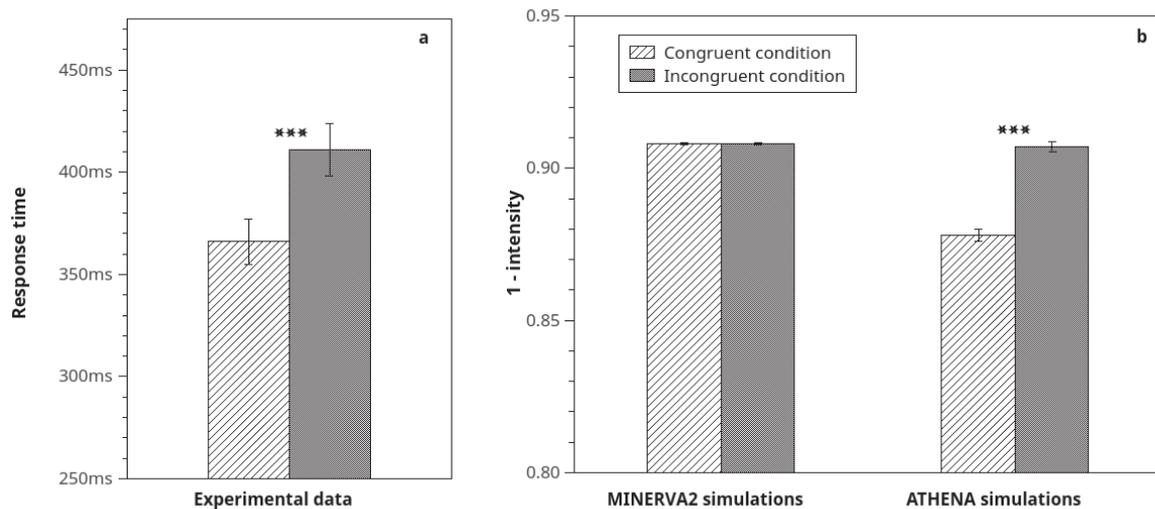


Figure 3: Heurley & al. (2012). Congruence effect on response from experimental (a) and simulations (b) data

traces and between these traces and the situation. Likewise, for a Stimulus-Response Compatibility (SRC) effect to emerge, it was necessary to adapt former sensorimotor knowledge (i.e. about the way we grasp objects) to the situation. As a consequence, both effects could only emerge and interact if memory traces interactions were interacting in a totally dynamical way.

Firstly, this experiment demonstrated that the memory system needs to be able to produce transitive knowledge in order to respond to a given cue. In this experiment, participants learned that each name in the study phase was *old*. They also learned that in order to respond *old*, they had to perform an action. Finally, they already knew that each name had a visual representation as an object. Thus, responding *old* to an image involved the emergence of transitive knowledge from the image to the action used to respond. The same logic could be applied for the stimulus-response compatibility effect.

Secondly, this experiment also demonstrated, throughout the additive pattern of the old/new effect and the SRC effect, that knowledge emergence cascades. Indeed, the pattern of results could only be explained if processing the cue involved emergence about its status (i.e. *old* or *new*) and the way we grasp this object, and then these two emerging knowledges allowed the emergence of the response. In that case, the memory system needed to be able to process emerging knowledge in the same way it processed traces.

Simulation procedure. The simulation of this ex-

periment in ATHENA and MINERVA2 went through three phases: an initialization phase, a learning phase and a recognition phase. The initialization phase was designed as a set of situations allowing the models to acquire the necessary general knowledge for the processing of the experiment, which included the correlation between an image and a name, and the correlation between a graspable part orientation and the proper hand used to interact with this part (e.g. handle on the left correlated in the left hand). For this simulation, the instruction had to be correlated in the learning phase to answer *old* or *new* to an item with the appropriate hand. Thus, the learning phase presented to the models a list of object names to be learned correlated to the instruction (e.g. an *old* vector and a right hand vector). Which, through contextual negation due to the use of negative probabilities, made it possible to answer *new* to the perception of a new item. Finally, a recognition task presented to the models the image of an object with an oriented graspable part in order to observe their response—i.e. whether they recognized the object as *old* or *new*. For each of these phases, the situations within were presented in a random order.

The models only deal with vectors of numbers, so we attributed several features to each component used. In order to avoid a deterministic conception of this arbitrary attribution, as well as possible mathematical biases, to each component a random number of features were attributed, randomly placed in the vector and containing random values $\in \{-1, 1\}$ in accordance with

the Hintzman protocol (1986, 1988). These attributions were used only for one experimental simulation (equivalent to one participant), and then were randomly re-assigned 5,000 times for the other experimental simulations to generate statistical validity (which yields the equivalent of 5,000 participants for each type of response observed). The necessary components assigned were the name of an object, the graspable part orientation (e.g. handle to the right or to the left), the hand used (right or left) and the motor response (*old* or *new*); these components were created according to the procedure used in simulation 2.

Predictions. To run a simulation of this experiment on MINERVA2 and ATHENA, we used only low-level sensorimotor data. It is important to stress that both the motor response used to perform the recognition judgment and the motor response depending on the orientation of the object are emerging concepts. Thus a simulation of this experiment would only succeed if the models were capable of manipulating these emerging concepts (just like they can manipulate memory traces). The fractal activity of the ATHENA model should provide the similarity between process and content necessary for this dynamical adaptation (thereby processing some processes), while the MINERVA2 model should not be able to do so from its standard mnemonic processing.

First of all, the task in itself required the emergence of transitive knowledge. Analytically, the presentation of an image as a cue during recognition (A) should activate the name of the object (B), which in turn should activate the instruction for the response (C). However, although only the correlations between A and B (object knowledge) and between B and C (learning phase) were instantiated, it would be necessary to carry out a correlation of correlations in order to respond. This transitive emergence is syllogistic: if $A = B$ and $B = C$, then $A = C$. The presentation of the image should be enough to make the motor response emerge. However, we are not in a procedural approach to memory: the process should not be linear, but global and dynamic. This optimal process should emerge from the global matching of the combined activity of the traces as a whole. For the MINERVA2 and ATHENA models, observing a response behavior would demonstrate that traces are dynamically processed, whereas the absence of a response would reveal that traces are independently processed.

Secondly, the presentation of the image should not only act as a reminder of the name of the object seen, but also of an orientation process corresponding to the

SRC effect. Analytically, the position of a handle (A) should activate the usable hand (B), which should modify the response that the subject gives with this hand (C). Here we find the same transitive principle as above: for the given response to be influenced by the orientation of the handle, it would be necessary to combine the memories linked to the grasping of a handle with the memories linked to the task instruction. As in the first point, this processing operation would not be procedural but contextual.

Thirdly, an interaction between these two types of emergence was observed in the original experiment: a compatibility of these emergences facilitated the construction of the memory, while an incompatibility impeded it. The response pattern therefore demonstrated an emergence of interactions between these two emerging concepts: a cascade emergence. To show that a model would be capable of handling the two emerging concepts defined above, we should find in the simulation results an interaction pattern consistent with the experimental results.

In order to be able to process its own abstraction emergence, MINERVA2 would need to re-inject its echo as a probe. Since this simulation would require cascade emergences, this model would only be able to simulate this experiment if the echo values were re-injected as a new probe several times in a row. But each transitive emergence needs a new echo re-injection in order to be processed: an external intervention has to choose how many re-injections are needed for the experiment. However, this is not the case for a human participant. To simulate the mnemonic peculiarities of this experiment more ecologically, only a single probe should be necessary in order to simulate efficiently this experiment. ATHENA should produce as many cascade emergences as needed through one probe injection without any external help.

Results analysis. The responses of the models were observed in the echo corresponding to the questions asked during the recognition phase—i.e. equations (3) and (12). Indeed, the echo corresponds to the totality of the model's response event (motor functions as well as the cognitive activity). In the same way that only the motor response of participants is used in laboratory experiments to analyze their psychological processes, we looked only at the numerical part of the echo C_j associated with the motor response of the models to analyze their processing operations. Thus, we did not seek merely a modification of the response time, but a modification in the reconstruction of memory. Therefore, we can observe in Figure 4 the behavioral results of

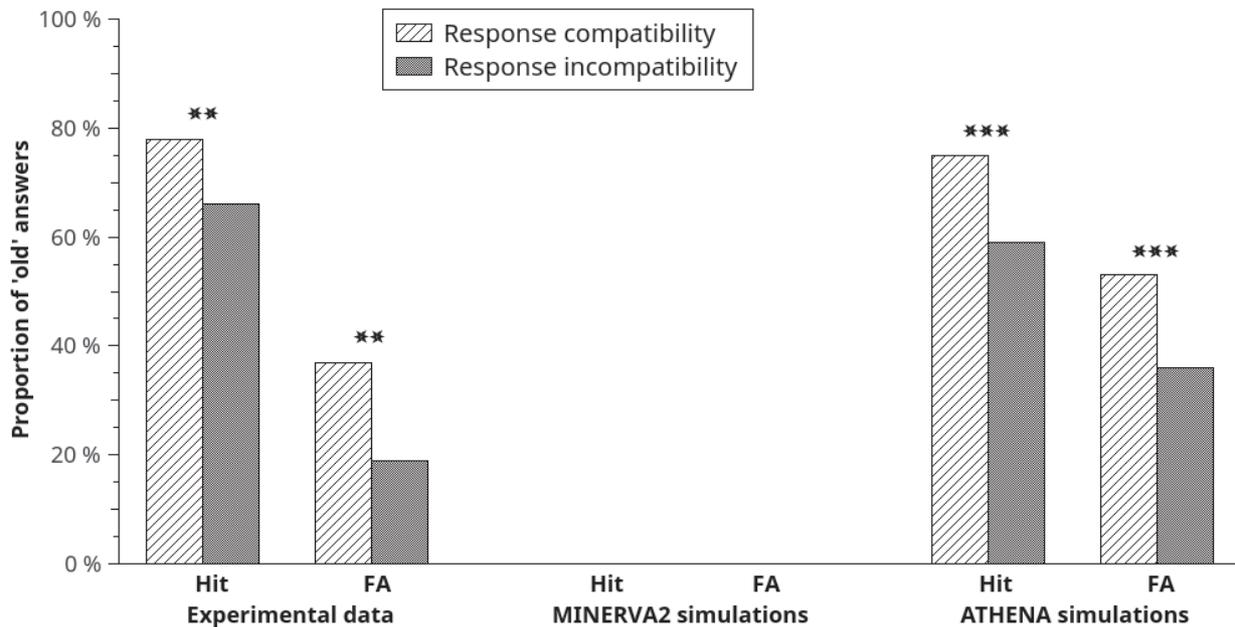


Figure 4: Brouillet & al (2015). Proportions of Hit and False Alarm according to compatibility from experimental and simulations data

this experiment and the simulations carried out by MINERVA2 and ATHENA. After normalizing the sign, the response characteristic of the model is a float number, where a negative value corresponds to a *new* motor response signifying that the model judged the item as *new*, and where a positive value corresponds to an *old* motor response, signifying that the model judged the item as old. A value of zero is an absence of response, signifying that the model could not make a judgment. These results were analyzed by focusing on the proportion of *old* responses.

First of all, we can notice that MINERVA2 never produced any motor response (i.e. the collected values were always at 0). Indeed, as we saw in the second simulation, given that this model could not remember how it processed information, it was capable neither of performing dynamical traces activations nor of managing transitive emergence. Therefore the two first types of transitive emergence expected (the motor response and the SRC effect) did not emerge and the model never gave a motor response (the echo of the response features always remained at 0).

As expected, the pattern of responses of the ATHENA model was similar to the experimental data. The mere fact that the model gave motor responses allowed us to ensure that it was capable of simulating the first type

of syllogistic emergence (the adaptation of the memory to the instruction). We saw that the proportion of *old* responses was greater for old items (hit: 67 per cent) than for new items (false alarm: 44 per cent): $\chi^2(1) = 1026.070$, $p < 0.001$. Indeed, the model used the memory of the object with the response instruction, and it was able to respond accordingly.

The graspable part orientation of the item influenced the response given based on its compatibility with the hand used to answer. The percentage of *old* responses was greater in compatible (64 per cent) than in incompatible situations (47 per cent): $\chi^2(1) = 551.798$, $p < 0.001$.

Finally, we saw that for *old* items, the hit percentage was greater in compatible (75 per cent) than in incompatible (59 per cent) situations: $\chi^2(1) = 289.462$, $p < 0.001$. Likewise, for *new* items the false alarm percentage was greater in case of compatibility (53 per cent) than in case of incompatibility (36 per cent): $\chi^2(1) = 292.54$, $p < 0.001$.

These results demonstrated that the ongoing processes of the ATHENA model were not simply stored and reused, but that they were able to compete just like any set of traces. The ATHENA model compared two contextual emergences as if they were retrievable content. Global matching took place for the different traces

(as in MINERVA2), but also for the different processes that emerged from this global matching.

Each trace of the ATHENA model is directly linked to the others, whether by modifying the preceding traces that made it emerge, or by modifying itself as the creation of new traces is taking place. Therefore, the activation of a trace will necessarily provoke an activation of the other traces according to their contextual relevance. These multiple emergences will therefore be in dynamic interaction and compete to provoke the emergence of the optimal contextual interpretation of the whole memory. Enactivism can account for abstractions manipulation.

5. General Discussion

Abstraction manipulation is usually referred to as an argument in disfavor of enactivism. If sensorimotor processing can easily be described in enactivist terms, the complexity required for sensorimotor interactions to account for emergent abstraction manipulation can seem counter-intuitive. In order to put enactivism to the test in abstraction simulations, we developed a mathematical formalization of an enactivist memory model. In order to do so, ATHENA is built according to four theoretical prerequisites: synergistic ephory, sensorimotor covariances, contextualization by inference, and fractal recursions.

Synergistic ephory is the ability to construct a memory according to the global matching between the current situation and episodic traces. MINERVA2 is the mathematical formalization of the synergistic ephory principle, and we constructed ATHENA from an alternate version of MINERVA2's law of total probability.

Sensorimotor covariances are the enactivism's response to the *hard problem of content*. If memory does not learn retrievable mnemonic content, it should instead learn reusable mnemonic processes about encountered correlations. ATHENA does not learn the probe (nor the echo), but the correlation between a situation and the way information has been processed (the average values between probe and echo).

Contextualization by inference is the way enactivism can process the previously learned sensorimotor covariances. According to Havelange et al. (2003), what we generally call a representation is more accurately a re-presentation: a *presentification* of the past, the activity of being presented again. In this sense, a global-matching memory just needs to be able to contextualize itself to a new situation according to previous situations encountered. Each learned covariance has to try to interpret the current situation according to its specific-

ties: this is a contextualization process. From the global matching between these contextualizations can emerge an adapted re-presentation. To this extent, even a very concrete concept is an abstraction of past interpretations re-presented (such as the color red); therefore, differentiating low-level and high-level processing is just a matter of mathematical complexity.

Fractal recursions is what permits an enactivist memory to match several contextualization processes between them in order to manipulate abstractions as if they were mere content. The dissociation between process and content comes from the computo-symbolic paradigm. However, sensorimotor covariance is not content in the symbolic sense of the term, but a process (in the same way that chromatic perception or motricity are processes). Since we cannot compare several sensorimotor covariances as symbolic contents, we can only observe their interactions: their correlation information. In the same way that memorizing a correlation between an individual and its environment allows enactivism to reproduce a similar pattern in a similar situation, the correlation between the different covariances can be adapted to a similar cognitive situation. The cognitive system must learn how it processes some processes. Henceforth, designing an enactivist memory model requires the use of recursive covariances. Each learned sensorimotor covariance behaves like memory in its totality: it learns continuously according to the way in which it can process the information. Finally, an enactivist memory process is scale invariant. Since an enactivist memory requires the use of processes processing processes, there is a mathematical object able to adapt to our scale invariance requirements: the fractal pattern, which has been observed in the mathematical formalization of ATHENA.

These four theoretical prerequisites to model an enactivist memory has permitted us to propose a mathematical formalization of enactivism. ATHENA has been put to the test to ensure that enactivism is indeed able to perform abstraction emergence, abstraction learning and abstraction manipulation.

With each simulation, we made comparisons between the behaviors of MINERVA2 and its enactivist equivalent: ATHENA. The first simulation aimed to show that the fundamental abstraction emergence processing of MINERVA2 was still present in ATHENA even without retrievable content to abstract (reusable sensorimotor covariances can perform an abstraction). The DRM paradigm demonstrates that memory categorization is an emerging construction of global matching. Its simulation allowed us to ensure that the basic categorization

capacities of MINERVA2 were still possible with an enactivist model.

The second simulation aimed at testing both models' ability to keep trace of their ongoing memory processes. The Heurley et al. (2013) experiment demonstrated that the processing operation carried out during the presentation of the word played the role of a prime and had an influence on the response. This simulation showed that ATHENA was capable of learning how the information was processed (i.e. to learn emergent abstractions as if they were retrievable content), while MINERVA2 was not.

The third simulation intended to test both models' capacity to combine ongoing emergent processes, just as they can combine traces. The Brouillet et al. (2015) experiment showed that an ongoing sensorimotor simulation was a process able to modify a behavior. The Stimulus-Response Compatibility (SRC) effect and the contextual adaptation of the instructions were both emerging transitive knowledge in competition. By combining process and content, the fractal characteristics of the ATHENA model allowed it to operate a global matching process with concepts emerging from the basic global matching between traces. This model can compare and manipulate emerging knowledge as efficiently as traces. MINERVA2 did not show this ability without an exterior intervention.

The simulations carried out validated the principle underlying ATHENA. This model does not memorize retrievable content, but the memory processes themselves in order to re-use them *in situ*. Its originality resides in the fact that it put to the test the enactivist theory whereby sensorimotor covariances can account for abstraction emergence, abstraction learning and abstraction processing. In accordance with Act-In theories, trace dynamics are sufficient if they learn only how they process information: the meanings of these processes are a contextual construction. The contextualization of fractal covariances allows ATHENA to manipulate emergent abstractions. Therefore, ATHENA is an improvement over single-system multiple-trace modeling. Where MINERVA2 was only able to account for the emergence of memories by synergistic ephory (Tulving, 1976, 1982, 1983), ATHENA is able to account for the historicity of memories, and thus its own dynamic. It is the fractal functioning of global matching that allowed us to account for the contextual dynamics of memory traces—their constructed and reconstructed characteristics known since Bartlett (1933).

The ATHENA model emphasizes that enactivism can use sensorimotor covariances to account for the con-

struction of abstraction. In addition, understanding the contents of memory as the memory of processes not only allows us to account for all memories, whatever their level of abstraction, but also to adapt any memory to any new situation. Enactivism can entirely process complex abstractions.

6. Future works

The fractal nature of ATHENA traces and the constant use of the entire range of memory X as prior knowledge cause an exponential increase in computational power for each new situation encountered. This was not a problem for the simulations presented in this study, but in order to simulate longitudinal studies, we would be limited by computing power. The sensorimotor nature of this enactivist model would make neural networks a good candidate for a more ecological implementation. We are therefore considering modeling the ATHENA equations in a neural network. This way, the necessary computing power would remain constant over time, and we could test it with more complex learning across a longer-term perspective. A neural network with the capacity to manipulate its own emerging abstractions would be able to adapt one context to the other (e.g. for an automated contextualization in translation) and be able to learn how to learn.

When ATHENA was being developed, we observed in it a striking temporal construction. In the first place, we observed a very prominent serial position effect, which required us, in this study, to control the presentation order of the probes by making it random. However, no temporal distinction is memorized during learning or recognition. Could the mere ability to learn the way information is processed be enough to account for the serial position effect? Franklin and Mewhort (2015) have been able to capture the serial position curve in a single-system memory (without a short-term memory nor a long-term memory implemented). The non-independent characteristics of memory traces may be the common trait shared by their model and ATHENA, able to account to such a phenomenon. So it will be necessary to conduct further analysis.

Moreover, we observed a memory prediction phenomenon similar to a learned temporal pattern. The same phenomenon can be observed when a memorized phone number is recalled (recalling a number activates the following ones, it may even be impossible to recall the last numbers until the first ones are recalled). In ATHENA, the echo accounts not only for the interpretation of the memory in the present situation, but also

for processes that was near in time to the coding of this emerging memory. How could a fractal memory model be enough to account for time prediction? A forthcoming study will examine how these specific time particularities of the ATHENA model can emerge from the mere dynamical nature of an enactivist memory.

References

- Arndt, J. (2010). The role of memory activation in creating false memories of encoding context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1):66.
- Atkinson, R. C. and Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of learning and motivation*, 2:89–195.
- Bartlett, F. C. and Burt, C. (1933). Remembering: A study in experimental and social psychology. *British Journal of Educational Psychology*, 3(2):187–192.
- Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *The Journal of General Psychology*, 49(2):229–240.
- Bransford, J. D. and Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive psychology*, 2(4):331–350.
- Brouillet, D., Brouillet, T., Milhau, A., Heurley, L., Vagnot, C., and Brunel, L. (2015). Word-to-picture recognition is a function of motor components mappings at the stage of retrieval. *International Journal of Psychology*, 51(5):397–402.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of experimental psychology*, 58(1):17.
- Feynman, R. P. (1987). Negative probability. In Hiley, B. and Peat, D., editors, *Quantum implications: essays in honour of David Bohm*, pages 235–248. Routledge & Kegan Paul London, UK.
- Franklin, D. R. and Mewhort, D. (2015). Memory as a hologram: an analysis of learning and recall. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 69(1):115.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138.
- Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, 100(1):70–87.
- Havelange, V., Lenay, C., and Stewart, J. (2003). Les représentations: mémoire externe et objets techniques. *Intellectica*, 35:115–131.
- Heurley, L. P., Brouillet, T., Chesnoy, G., and Brouillet, D. (2013). Color perception involves color representations firstly at a semantic level and then at a lexical level. *Cognitive processing*, 14(1):19–29.
- Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review*, 93(4):411–428.
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95(4):528.
- Hutto, D. and Myin, E. (2012). *Radicalizing enactivism: Basic minds without content*. MIT Press.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological review*, 95(4):492.
- Maturana, H. R. and Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding*. New Science Library/Shambhala Publications.
- O’Regan, J. K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24(5):939–973.
- Piaget, J. (1936). *La naissance de l’intelligence chez l’enfant*. Delachaux.
- Piaget, J. (1937). *La construction du réel chez l’enfant*. Delachaux & Niestlé.
- Ratcliff, R. and McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological review*, 95(3):385.
- Roediger, H. L. and McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of experimental psychology: Learning, Memory, and Cognition*, 21(4):803.
- Roediger, H. L., Rajaram, S., and Srinivas, K. (1990). Specifying criteria for postulating memory systems. *Annals of the New York Academy of Sciences*, 608(1):572–595.

- Rousset, S. (2000). Les conceptions "système unique" de la mémoire: aspects théoriques. *Revue de neuropsychologie*, 10(1):27–51.
- Tucker, M. and Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human perception and performance*, 24(3):830.
- Tulving, E. (1976). Ecphoric processes in recall and recognition. In Brown, J., editor, *Recall and recognition*. John Wiley & Sons.
- Tulving, E. (1982). Synergistic ecphory in recall and recognition. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 36(2):130.
- Tulving, E. (1995). Organization of memory: Quo vadis. *The cognitive neurosciences*, pages 839–847.
- Tulving, E., Le Voi, M., Routh, D., and Loftus, E. (1983). Ecphoric processes in episodic memory [and discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 302(1110):361–371.
- Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of consciousness studies*, 3(4):330–349.
- Varela, F. J., Thompson, E., Rosch, E., et al. (1993). L'inscription corporelle de l'esprit. *Paris, Seuil*.
- Versace, R., Labeye, E., Badard, G., and Rose, M. (2009). The contents of long-term memory and the emergence of knowledge. *European journal of cognitive psychology*, 21(4):522–560.
- Versace, R., Vallet, G. T., Riou, B., Lesourd, M., Labeye, É., and Brunel, L. (2014). Act-in: An integrated view of memory mechanisms. *Journal of Cognitive Psychology*, 26(3):280–306.
- Whittlesea, B. W. (1987). Preservation of specific experiences in the representation of general knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(1):3.
- Whittlesea, B. W. (1989). Selective attention, variable processing, and distributed representation: Preserving particular experiences of general structures. In Morris, R., editor, *Parallel distributed processing: implications for psychology and neurobiology*, pages 76–101. Clarendon Press/Oxford University Press.
- Whittlesea, B. W. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(6):1235.
- Whittlesea, B. W. (1997). Production, evaluation, and preservation of experiences: Constructive processing in remembering and performance tasks. *The Psychology of Learning and Motivation*, 37:211–264.

Simulations

Les simulations présentées dans l'article furent nécessaire pour s'assurer qu'un modèle énaïviste était capable de produire, d'apprendre et de manipuler des abstractions émergentes sans employer de représentation. Cependant, celles-ci ne sont suffisantes ni pour s'assurer que ATHENA reproduit les mêmes capacités mnésiques que MINERVA2, ni pour mettre en exergue ce qui différencie les deux modèles dans leur traitement de l'information. Pour y remédier nous allons présenter deux séries de simulations : une première relative au fonctionnement normal de MINERVA2, ainsi qu'une seconde relative à un comportement temporel émergent prédit par l'inter-dépendance des traces d'un modèle énaïviste.

3.1 ATHENA et MINERVA2

La première simulation (i.e., paradigme DRM) que nous avons présentée dans l'article du chapitre précédent a permis de s'assurer que le modèle ATHENA est capable de reproduire les capacités du modèle MINERVA2 qui, rappelons-le, est la formalisation mathématique du processus d'écphorie synergétique (appariement global). Mais pour pouvoir affirmer que notre modèle est aussi efficace que son prédécesseur, il est nécessaire de reproduire les autres apports de MINERVA2 sur la modélisation de la mémoire. Nous allons ici présenter quatre simulations de phénomènes mnésiques qui peuvent être reproduits par MINERVA2 avec les résultats conjoints des deux modèles pour pouvoir les comparer.

La première simulation (Rappel Indicé) a pour but de s'assurer que ATHENA parvient à reproduire deux comportements mnésiques fondamentaux : la récupération d'une information à partir de données incomplètes ainsi que l'amélioration de l'apprentissage par répétition. L'objectif de la deuxième simulation (FURIG et Prototypie) est de montrer que ATHENA reproduit les particularités d'une mémoire épisodique à système unique. La troisième simulation (FURIG et Réaction à la nouveauté) vient compléter la deuxième simulation en s'assurant que la nouveauté est plus surprenante que ce que la prototypie est familière. La quatrième simulation (MCE) complète la simulation du paradigme DRM en s'assurant qu'au delà des faux souvenirs propres à une recombinaison intra-composante, ATHENA est également capable de produire des faux souvenirs propres à une recombinaison inter-composantes.

3.1.1 Rappel indicé

Cette simulation reprend les capacités basiques attendues d'un modèle de mémoire : (i) la mémoire doit permettre l'émergence d'un souvenir complet à partir d'informations incomplètes, et (ii) la capacité de faire émerger un souvenir doit s'améliorer avec la répétition d'informations.

Simulation originelle. Dans le premier article où est exposé le modèle MINERVA2, Hintzman (1984) présente la simulation du rappel indicé de manière simplifiée : lors de la perception d'un visage, la mémoire permet de se souvenir du nom de la personne (bilatéralement, il est aussi possible de se souvenir d'un visage à l'évocation d'un nom). Cette récupération devient de plus en plus efficace à mesure que nous sommes confrontés à l'association visage-nom d'un individu. Pour cette simulation, il faut présenter au modèle plusieurs associations visage-nom différentes afin qu'il puisse démontrer sa capacité à choisir l'information pertinente en fonction du visage perçu. La simulation sera reproduite avec plusieurs présentations de la même association visage-nom afin de tester l'influence de la répétition sur l'efficacité du souvenir.

Procédure de simulation. Pour cette simulation, Hintzman a proposé l'emploi de 20 composantes : les dix premières étant réservées aux informations liées aux visages (informations visuelles), les dix dernières aux informations liées aux noms (informations auditives). Aucune distinction n'est faite pour délimiter les catégories de composantes, elles sont censées émerger en situation. Ainsi, aussi bien pour MINERVA que pour ATHENA, les deux modèles peuvent apprendre une association visage-nom en leur présentant une sonde où les 20 composantes ont une activité signifiante (activations et inhibitions aléatoires mais stables dans l'activité qui les lie aux caractéristiques visuelles et auditives). Par exemple une première association visage_A (présenté en gras pour faciliter la lecture) et nom_A (non gras) sera la sonde (**1, -1, -1, 1, -1, -1, 1, 1, 1, -1, -1, -1, 1, 1, -1, 1, 1, 1**). La présentation d'un visage seul s'effectue à l'aide d'une sonde où les composantes visuelles ont une activité signifiante, ainsi qu'une activité nulle pour les composantes auditives. Par exemple, le visage_A sans le nom sera la sonde (**1, -1, -1, 1, -1, -1, 1, 1, 1, -1, 0, 0, 0, 0, 0, 0, 0, 0**).

Nous employons ici une instance de chaque modèle comme un sujet auquel nous faisons passer une expérience. Chaque simulation est reproduite 5000 fois pour pouvoir analyser statistiquement les résultats comme s'ils provenaient d'autant de participants. Dans la tâche d'apprentissage, chaque sujet apprend une liste de plusieurs associations visage-nom. Dans cette liste il y a trois associations aléatoires de remplissage qui ne sont pas l'objet de l'expérience, et une association cible que le sujet doit être capable de rappeler. Il y a cinq groupes de sujets pour lesquels l'association cible est présentée 0, 1, 2, 3 ou 4 fois dans la liste d'apprentissage. Dans la tâche de rappel, nous présentons aux sujets le visage de l'association cible et nous observons dans l'écho le nom rappelé.

Hintzman n'a pas présenté les résultats de cette expérience, certainement parce que l'aspect aléatoire des sondes complique l'analyse des résultats de l'écho. Pour faire passer l'expérience aux deux modèles, nous avons dû mettre au point une méthode d'évaluation de la réponse pour quantifier la qualité de réponse des modèles. Celle-ci évalue la distance moyenne inversée et recentrée entre les composantes des vecteurs étudiés (vecteur attendu et écho produit) par l'équation $(\frac{-1}{m} \sum_{j=1}^m |x_j - y_j|) + 1$ (où x_j est la composante j du premier vecteur, y_j est la composante j du second vecteur, et m le nombre total de composantes). Nous sommes ainsi en mesure de comparer les résultats des sujets entre eux. Cette équation donne une valeur comprise entre -1 (pour deux vecteurs opposés) et 1 (pour deux vecteurs identiques). En comparant des vecteurs aléatoires, l'espérance de cette équation est de 0.33 (et non pas 0 car il existe statistiquement plus de sondes semblables que différentes), et son écart-type est de 0.13. Cette équation prend alors une marge d'erreur entre 0.20 et 0.46 où la distinction entre vecteurs semblables et différents est floue (ce qui permet une meilleure relativité des résultats).

Prédictions. Cette simulation doit donner des comportements similaires pour les deux modèles : ils doivent être capables de retrouver efficacement le nom associé au visage cible, et ce de manière plus efficiente avec la répétition perçue des associations cibles. Si apprendre des covariances entre la perception et la production mnésique plutôt que des représentations de visages et de noms est aussi efficace que le prétend l'énactivisme radical, le modèle ATHENA doit être capable de reproduire les capacités du modèle MINERVA2.

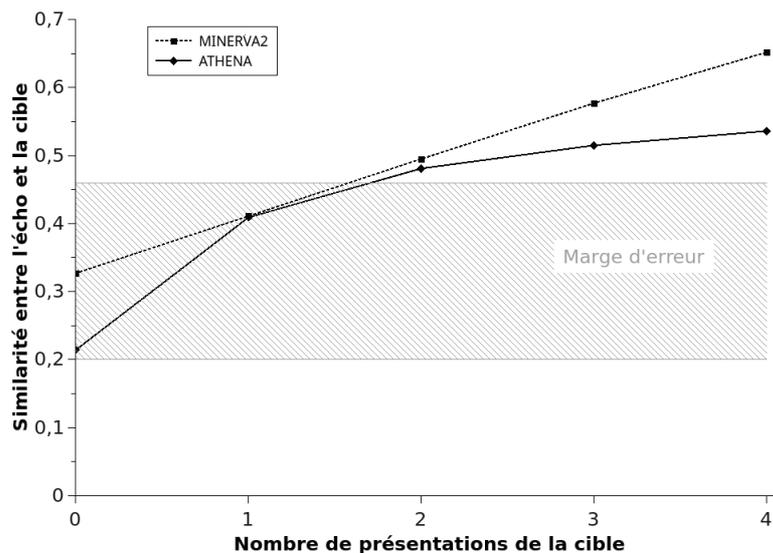


FIGURE 5 – *Rappel indicé* : Similarité entre écho et cible selon le nombre de présentations de la cible pour MINERVA2 et ATHENA

Résultats. En premier lieu, nous pouvons voir dans la figure 5 que lorsqu'il n'y a

eu aucune occurrence de l'association cible dans la phase d'apprentissage, le modèle MINERVA2 répond dans l'écho un nom dont la distance au nom cible (0.33(0.14)) correspond à l'espérance de la fonction de comparaison. Ceci signifie que le modèle ne peut pas rappeler un nom qu'il ne connaît pas et que sa réponse est aléatoire. Le modèle ATHENA, quant à lui, n'est pas capable de rappeler un nom qu'il n'a pas appris non plus, mais sa réponse se différencie plus du nom cible que ce que donnerait une réponse aléatoire (0.21(0.16)). La différence de valeur entre les deux modèles est significative ($T(9998) = 39.465$; $p < 0.001$). Cette particularité est due au fait que le modèle ATHENA tente de reconstruire un souvenir à partir de ce qu'il connaît déjà (les associations apprises d'autres visages et d'autres noms), ce qui diffère forcément plus du nom cible qu'une réponse aléatoire. Tandis que MINERVA2 compare la sonde avec du contenu représentationnel (des sondes apprises), et comme la nouvelle sonde est différente des sondes précédentes de manière aléatoire, la différenciation que ce modèle en fait suit la même loi statistique aléatoire. ATHENA tente de reconstruire un souvenir là où MINERVA2 tente de le récupérer.

En second lieu, les courbes des deux modèles augmentent au fur et à mesure du nombre de fois auquel ils ont été confrontés à l'association cible. Le modèle MINERVA2 suit une courbe longiligne et son évolution est statistiquement significative comme le montre un test de Pearson à $r(25000) = 0.556$; $p < 0.001$. Chacun des groupes a un résultat différent du suivant, et la présentation s'améliore à chaque présentation supplémentaire de l'association cible :

Test de Student entre les positions 1 et 2 : $T(9998) = 27.224$; $p < 0.001$

Test de Student entre les positions 2 et 3 : $T(9998) = 25.231$; $p < 0.001$

Test de Student entre les positions 3 et 4 : $T(9998) = 23.02$; $p < 0.001$

Test de Student entre les positions 4 et 5 : $T(9998) = 19.901$; $p < 0.001$

L'évolution du modèle ATHENA est tout autant significative avec un test de Pearson à $r(25000) = 0.575$; $p < 0.001$. La distinction entre chaque groupe est aussi significative :

Test de Student entre les positions 1 et 2 : $T(9998) = 63.759$; $p < 0.001$

Test de Student entre les positions 2 et 3 : $T(9998) = 26.846$; $p < 0.001$

Test de Student entre les positions 3 et 4 : $T(9998) = 12.895$; $p < 0.001$

Test de Student entre les positions 4 et 5 : $T(9998) = 7.527$; $p < 0.001$

Enfin, il convient de remarquer que si les résultats sont conformes aux attentes pour les deux modèles, leurs courbes n'ont pas la même forme. La courbe de MINERVA2 est une droite où chaque nouvelle occurrence supplémentaire apprise améliore les résultats de la même manière. La courbe de ATHENA suit une progression logarithmique où chaque nouvelle occurrence supplémentaire apprise améliore les résultats relativement à ce que le modèle serait déjà capable de faire (la progression est rapide pour les premières occurrences et diminue selon le nombre de répétitions effectuées). Cette courbe d'apprentissage est typique de l'expérience humaine (Ebbinghaus, 1913; Bills, 1934) et est due au fait que la présentation d'un nouvel item améliore son souvenir relativement au nombre de fois

où l’item a déjà été présenté. Un item non fluent a besoin de moins de répétitions pour augmenter sa fluence comparativement à un item déjà connu et fluent. Conformément à l’apprentissage humain, l’apprentissage du modèle ATHENA suit une logique de contextualisation avec une normalisation des informations apprises. Ce que l’on peut observer dans les résultats de cette simulation par la forme logarithmique de la courbe d’apprentissage du modèle.

Cette simulation nous a permis de nous assurer que le rappel indicé et l’amélioration par répétitions ne nécessitent pas d’apprendre des représentations pour pouvoir être efficace, mais aussi que la courbe d’apprentissage est plus pertinente pour un modèle énonciviste que pour son équivalent représentationnaliste. En effet, pour un sujet humain l’apprentissage est logarithmique et non pas linéaire.

3.1.2 FURIG et Prototypie

Le rappel indicé est un comportement typique de la mémoire mais n’est pas propre à un paradigme en particulier. Puisque MINERVA2 et ATHENA sont avant tout des modèles de mémoire épisodique, il est nécessaire de les confronter à une expérience typique de la mémoire épisodique. L’expérience dénommée « FURIG » (Whittlesea, 1987) a démontré que la mémoire humaine ne se contente pas d’apprendre une abstraction des situations rencontrées, mais les situations elles-mêmes. Par souci de cohérence, un modèle épisodique de type MINERVA2 doit être capable de reproduire ces résultats.

Expérience originelle. Le matériel de l’expérience de Whittlesea (1987) consiste en la génération de trois listes de cinq pseudo-mots de cinq lettres (de type C-V-C-V-C) construits en modifiant deux des cinq lettres d’un pseudo-mot prototypique (e.g., FURIG). La première liste est générée en ne respectant que cette règle (e.g., FEKIG, FUTEG, PURYG, FYRIP, KURIT). La deuxième liste respecte le fait d’avoir des exemplaires différents de deux lettres du prototype, mais a aussi la particularité d’avoir des exemplaires différents d’une lettre de la première liste (pour l’item le plus proche : e.g., FYKIG, FUTYG, PUREG, FERIP, PURIT). La troisième liste a également deux lettres de différence avec le prototype, mais aussi deux lettres de différence avec la première liste (e.g., FUKIP, PUTIG, FURYT, FYREG, KERIG). Toutes les listes ont le même écart avec le prototype mais un écart différent entre elles.

Lors de la phase d’apprentissage, la première liste est présentée aux sujets, un exemplaire à la fois. Ceux-ci doivent écrire chaque exemplaire présenté, et l’exemplaire reste présent à l’écran jusqu’à ce que les sujets aient fini de l’écrire. Une fois que tous les exemplaires de la liste ont été retranscrits, une phase de test présente dans un ordre aléatoire chaque exemplaire des trois listes pendant 30ms chacun. Pour chaque item, les sujets ont pour consigne d’écrire le mot perçu (comme dans la phase d’apprentissage), mais sans que l’exemplaire ne reste affiché à l’écran. Un score quantifiant la qualité de retranscription

est attribué aux sujet pour chacune des trois listes.

Si la mémoire des sujets fait appel à une abstraction automatique des informations, ils ne devraient apprendre qu'un prototype des exemplaires (FURIG dans notre exemple). Dès lors, et comme les trois listes sont également éloignées du prototype, il ne devrait pas y avoir de différence de score entre les listes. Cependant, si la mémoire des sujets est épisodique, seuls les exemplaires de la première liste devraient être appris. La deuxième liste devrait donner un score plus faible que la première liste, mais surtout la troisième liste devrait donner un score plus faible que la deuxième liste (i.e., liste 1 > liste 2 > liste 3).

L'auteur a mis en lumière le fait que la distance aux exemplaires appris est inversement proportionnelle au score obtenu. Les sujets n'apprennent donc pas un prototype des informations apprises mais les exemplaires rencontrés, ce qui argumente en faveur de la théorie d'une mémoire épisodique à système unique.

Procédure de simulation. L'expérience originelle utilise trois lettres différentes pour chaque position de lettre dans le mot (e.g., F, P et K pour la première lettre). Dès lors, pour pouvoir simuler cette expérience avec des valeurs de composantes entières (1 et -1), il est nécessaire de faire appel à deux composantes par lettre (e.g., (1, 1), (-1, -1) et (-1, 1) pour F, P et K). Chaque pseudo-mot de cinq lettres est alors généré avec un vecteur de 10 composantes. Un prototype est choisi aléatoirement parmi les lettres disponibles et les trois listes sont générées en respectant les mêmes différences structurelles que celles employées par Whittlesea : les trois listes ont des exemplaires n'ayant que deux lettres de différence avec le prototype, la seconde liste contient des exemplaires éloignés d'une lettre par rapport à l'item le plus semblable de la première liste, et la troisième liste contient des exemplaires éloignés de deux lettres par rapport à l'item le plus semblable de la première liste. Chaque lettre est générée aléatoirement de manière à être unique en utilisant le protocole de similarité entre vecteurs décrit dans la simulation précédente : deux vecteurs sont considérés comme différents à partir du moment où leur valeur de similarité est inférieure à la marge d'erreur (i.e., $< 0,2$).

Lors de la phase d'apprentissage, les items de la première liste sont présentés aux modèles dans un ordre aléatoire. Puis lors de la phase de reconnaissance, les trois listes sont présentées aux modèles en mode test (i.e., sans que les modèles n'apprennent de nouvelles informations, de manière à s'assurer que la reconnaissance d'un mot n'influence pas la reconnaissance des autres). L'intensité du traitement de la reconnaissance est observée de manière à donner un score à l'efficacité de l'interprétation des sondes selon la liste étudiée. La simulation est reproduite 5000 fois pour les deux modèles, puis les intensités des trois listes sont comparées pour chaque modèle.

Prédictions. MINERVA2 et ATHENA sont tous deux des modèles de mémoire épisodique dont la capacité d'abstraction ne provient pas d'un processus d'apprentissage

prototypique mais d'un appariement global de leurs traces épisodiques mises en situation. Les deux modèles devraient obtenir les mêmes résultats que ceux obtenus dans l'expérience originelle et obtenir des scores différents aux trois listes (i.e., liste 1 > liste 2 > liste 3).

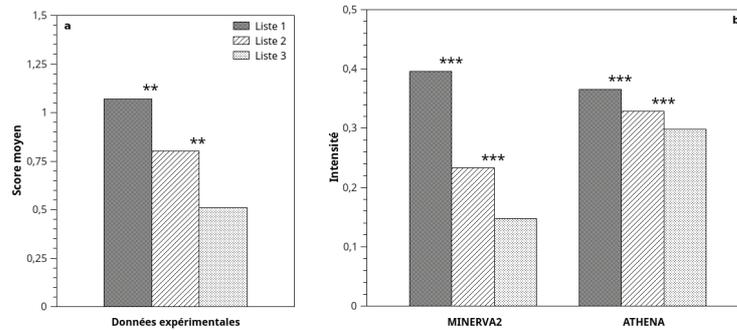


FIGURE 6 – *FURIG* : Scores et intensités moyens des trois listes pour les données expérimentales et les simulations des modèles MINERVA2 et ATHENA

Résultats. Nous pouvons remarquer en premier lieu que le pattern de résultats est le même pour les données expérimentales (Figure 6.a) et les simulations des deux modèles (Figure 6.b). Pour MINERVA2, l'intensité moyenne des exemplaires de la première liste (0.396(0.22)) est supérieure à l'intensité moyenne des exemplaires de la deuxième liste (0.233(0.16)) de manière significative ($T(49998) = 92.691$; $p < 0.001$). De même, l'intensité moyenne des exemplaires de la deuxième liste est supérieure à l'intensité moyenne des exemplaires de la troisième liste (0.148(0.11)) de manière significative ($T(49998) = 66.495$; $p < 0.001$). En ce qui concerne ATHENA, l'intensité moyenne des exemplaires de la première liste (0.365(0.16)) est supérieure à l'intensité moyenne des exemplaires de la deuxième liste (0.329(0.14)) de manière significative ($T(49998) = 25.838$; $p < 0.001$). De même, l'intensité moyenne des exemplaires de la deuxième liste est supérieure à l'intensité moyenne des exemplaires de la troisième liste (0.298(0.12)) de manière significative ($T(49998) = 26.641$; $p < 0.001$).

Sur le graphique, les résultats de MINERVA2 semblent plus marqués, mais cette particularité n'est due qu'à l'absence de bornes pour l'intensité de ce modèle (l'intensité est proportionnelle au nombre de traces). En effet, pour ATHENA l'intensité est bornée (l'intensité est comprise dans l'intervalle $[-1;1]$ quel que soit le nombre de traces) par une inférence de contextualisation qui semble atténuer les valeurs d'intensité. Cependant, les différenciations entre les valeurs d'intensité sont également contextualisées : les différences semblent plus faibles uniquement car une valeur proche des bornes (-1 et 1) est statistiquement plus difficile à atteindre que pour MINERVA2. Cette particularité ne remet pas en question les patterns de résultats obtenus par le modèle dans les simulations présentées dans ce travail.

Le caractère purement épisodique des traces de MINERVA2 et ATHENA leur per-

mettent d'obtenir des résultats conformes aux données expérimentales qui ne dépendent pas de la distance au prototype mais de la distance aux exemplaires. Nous pouvons tout de même remarquer que si MINERVA2 apprend directement les exemplaires qui lui ont été présentés, ATHENA n'apprend que la manière dont les exemplaires ont été traités. ATHENA reste plus en accord avec les travaux de Whittlesea (1997) stipulant qu'il n'existe pas de différence entre mémoire déclarative et procédurale, puisque toute forme d'information mémorisée se doit d'être procédurale pour pouvoir s'adapter à de nouvelles situations : on apprend toujours une manière de traiter les informations.

3.1.3 FURIG et Réaction à la nouveauté

Nous avons montré dans la simulation du paradigme DRM que les capacités d'abstraction de MINERVA2 se retrouvaient dans le modèle ATHENA à partir d'une mémoire épisodique. Cependant, ce dernier n'apprend pas directement des exemplaires pour construire une abstraction par synergie, il apprend la manière dont il serait capable de re-présenter ces exemplaires, dont il peut généraliser l'ensemble par inférence. Cette simulation a permis de s'assurer que l'abstraction n'est pas nécessairement une question de représentation, mais que la mise en commun de toutes les manières de traiter les informations passées permet de faire émerger une abstraction prototypique. La simulation de l'expérience de Whittlesea (1987) a exposé le fait que le prototype n'est pas un contenu récupéré mais un appariement construit. Cependant, ces deux expériences ne font que mettre en confrontation les expériences apprises avec le prototype de ces expériences, il ne s'agit toujours que d'abstraire une information conforme avec ce que les modèles ont appris. Mais qu'en est-il du traitement de nouvelles informations complètement différentes du prototype des expériences apprises ? L'énactivisme étant à l'origine une question d'adaptation à l'environnement, il serait intéressant de regarder comment les modèles essaient d'abstraire une information nouvelle et inattendue à partir des expériences passées (une nouvelle catégorie différente de celle du prototype). L'expérience de Whittlesea (1987) sur la prototypie nous semble être tout à fait indiquée pour tester la capacité des modèles à simuler le processus d'abstraction en œuvre quand nous sommes confrontés à des événements nouveaux, mais elle demande certains ajustements.

Expérience originelle. L'expérience de Whittlesea (1987) a pu montrer que l'abstraction par synergie des exemplaires épisodiques était plus efficace que par une dissociation entre mémoire épisodique et abstraction mnésique. Une liste de pseudo-mots différents de la liste d'apprentissage est moins bien reconnue que les exemplaires appris, mais qu'en est-il de la surprise provoquée par la présentation de cette nouvelle liste ? La différence entre familiarité et surprise se situe au niveau de la fluence d'un traitement (Whittlesea, 1997) : un traitement fluent provoque un sentiment de familiarité, tandis qu'un traitement non-fluent provoque un sentiment de surprise. Pour en rendre compte, une modification de

l'expérience originelle de 1987 a été proposée et simulée sur MINERVA2 (Brouillet & Jimenez, 2007) dans le but de quantifier la surprise provoquée par la présentation d'un mot nouveau et la comparer à l'intensité de familiarité. Et ce, afin de montrer que dans une mémoire d'appariement global la surprise est plus surprenante que ce que la familiarité est familière.

Dans cette simulation, une série de pseudo-mots est présentée comme liste d'apprentissage où chaque présentation est une légère modification d'un pseudo-mot prototypique non présenté. Par exemple, pour le pseudo-mot prototypique FURIG, la série présentée peut être : FYREP, PURIT, FEKIG, FUREG, FYRYG. Après la phase d'apprentissage, une phase de reconnaissance présente le prototype de cette liste (FURIG dans notre exemple) ainsi qu'un nouveau pseudo-mot issu de la modification du prototype (PUKIG dans notre exemple). Si le prototype FURIG est représenté par le vecteur $(1, 1, 1, 1, 1)$ où chaque composante du vecteur est une lettre, la liste de pseudo-mots de notre exemple peut être présentée à MINERVA2 sous la forme d'une liste de vecteurs telle que : $((1, -1, 1, -1, -1), (-1, 1, 1, 1, -1), (1, -1, -1, 1, 1), (1, 1, 1, -1, 1), (1, -1, 1, -1, 1))$. Lors de la phase de reconnaissance, les pseudo-mots FURIG $(1, 1, 1, 1, 1)$ et PUKIG $(-1, 1, -1, 1, 1)$ sont présentés au modèle et l'intensité de MINERVA2 est observée. Le prototype donne une intensité de 0.12 et le nouveau pseudo-mot de -1,22. La nouveauté est donc plus surprenante que ce que la prototypie est familière. Simuler cet exemple sur ATHENA donne des résultats similaires avec un prototype ayant une intensité de 0.22 et un nouveau pseudo-mot ayant une intensité de -0.40. Ce type de simulation donne des résultats intéressants, mais nous considérons que l'emploi de valeurs positives pour définir une lettre semblable au prototype (ainsi qu'une valeur négative pour définir ce qui doit être considéré comme nouveau) n'est pas suffisamment rigoureux. Ce problème ne peut pas être efficacement contrebalancé par l'emploi de sondes aléatoires, ce qui ne nous permet pas d'en faire des analyses statistiques. De plus, un pseudo-mot nouveau fait forcément partie d'une catégorie déjà étudiée par le sujet, ce qui limite l'étude des capacités d'adaptation à la nouveauté. Nous proposons alors une autre procédure de simulation utilisant deux catégories différentes pour remédier à ces problèmes.

Procédure de simulation. Pour pouvoir confronter les exemplaires de deux catégories différentes il est nécessaire d'utiliser deux types de distinction d'information (une différenciation entre les catégories et une différenciation entre les exemplaires d'une catégorie). Intuitivement nous pourrions proposer une différenciation inter-composantes et une différenciation intra-composante, mais MINERVA2 n'est pas capable de confronter les informations de traces ayant une distinction inter-composantes totale (i.e., sans qu'aucun lien ne soit établi entre les composantes ; e.g., les vecteurs $(1, -1, 0, 0)$ et $(0, 0, 1, 1)$ n'auront jamais aucun lien dans une simulation de MINERVA2). Ce modèle y parvient sans problème à la condition qu'une corrélation inter-composantes soit apprise (comme

pour les catégories nom/visage de la simulation du rappel indicé), mais nous cherchons justement à utiliser des catégories n'ayant aucune corrélation apprise. Cette particularité sera intéressante pour mettre les modèles à l'épreuve sur des critères temporels sans que d'autres interactions ne viennent perturber les résultats, mais elle n'est pas judicieuse pour la simulation que nous mettons en place dans le chapitre présent. Nous utiliserons dès lors une différenciation intra-composante de signes (positif et négatif) pour la génération de catégories, ainsi qu'une différenciation intra-composante de valeurs (comprises dans l'intervalle $[0;1]$) pour générer des exemplaires. Pour pouvoir générer aléatoirement des pseudo-mots capables de rendre compte de différences catégorielles, nous employons dans cette simulation des vecteurs ne contenant plus de simples valeurs entières, mais des valeurs flottantes. MINERVA2 n'est pas prévu pour analyser des vecteurs de valeurs flottantes, mais devrait pouvoir analyser efficacement la notion de catégorie. ATHENA est fait de manière à prendre en compte l'intensité relative des composantes de la sonde, nous allons ainsi pouvoir mettre cette particularité à l'épreuve à travers l'emploi de valeurs flottantes. De la même manière que dans une composante de vecteur l'activation (1) est différente de l'inhibition (-1), différentes valeurs flottantes de cette composante donnent des interprétations différentes (-0.5 est différent de 0.3, mais 0.9 est lui aussi différent de 0.3). En partant de ce principe, nous pouvons artificiellement créer plusieurs catégories en fonction des activations et inhibitions des composantes de vecteur, ainsi que plusieurs exemplaires dans ces catégories en jouant avec les valeurs flottantes de ces composantes. Par exemple, les vecteurs $(0.5, -0.8, 0.3)$ et $(0.2, -0.4, 0.8)$ sont des exemplaires d'un vecteur catégoriel composé d'une activation, d'une inhibition puis d'une activation. Le prototype de ces exemplaires en est la moyenne, soit $(0.35, -0.6, 0.55)$. Et un exemplaire d'une catégorie différente peut être $(-0.7, 0.6, 0.3)$, soit un vecteur composé d'une inhibition, et de deux activations.

La génération de sondes aléatoires demande de produire deux catégories différentes de valeurs entières (dans notre exemple : $(1, -1, 1)$ et $(-1, 1, 1)$), puis des valeurs flottantes aléatoires pour chaque exemplaire ayant le signe des composantes du vecteur de sa catégorie. Pour pouvoir générer aléatoirement des catégories différentes, nous utilisons la méthode de similarité des vecteurs développée dans la simulation du rappel indicé. Deux vecteurs aléatoires de valeurs entières sont générés jusqu'à ce que leur valeur de similarité soit inférieure à -0.6 (soit environ l'espérance de cette fonction à laquelle on soustrait sept fois l'écart-type : la valeur de -0.6 a été choisie pour avoir la valeur la plus basse possible sans trop ralentir les calculs de la simulation). Les exemplaires sont alors générés avec des valeurs flottantes aléatoires selon le signe de la catégorie correspondante : cinq exemplaires de cinq composantes pour la catégorie d'apprentissage ainsi qu'un exemplaire de cinq composantes pour la catégorie différente. Le prototype est calculé en faisant la moyenne de toutes les sondes d'apprentissage. Lors de la phase d'apprentissage, les cinq exemplaires de la catégorie d'apprentissage sont présentés aux modèles. Lors de la phase

de reconnaissance, les cinq exemplaires d'apprentissage, l'exemplaire de la catégorie différente ainsi que le prototype sont présentés aux modèles. La simulation est reproduite 5000 fois pour les deux modèles, et les intensités générées par ces sondes sont ensuite analysées.

Prédictions. Malgré l'emploi de valeurs flottantes, les signes des composantes catégorielles devraient permettre à MINERVA2 d'obtenir des résultats similaires à la simulation de Brouillet & Jimenez en confrontant le prototype à une nouvelle information. Mais l'absence de méthode propre aux valeurs flottantes ne devrait pas lui permettre de reproduire les résultats du paradigme DRM en comparant les résultats des exemplaires d'apprentissage au prototype. Ce n'est pas une limite directe du fonctionnement de MINERVA2 puisque ce modèle n'est pas prévu pour, mais cela nous permettra de nous assurer que la méthode d'utilisation de valeurs flottantes du modèle ATHENA parvient à généraliser le paradigme DRM dans son fonctionnement. En effet, ATHENA devrait non seulement être capable de reproduire le fonctionnement de MINERVA2 quant à la distinction prototype/nouveauté, mais aussi de conserver ses capacités d'abstraction quelles que soient les catégorisations possibles des données.

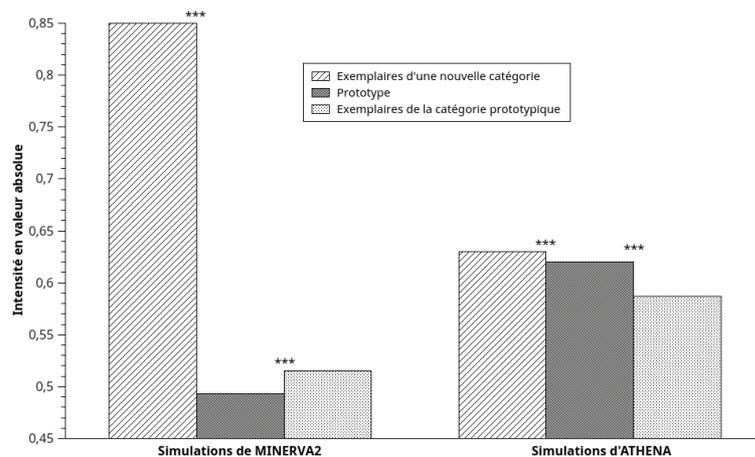


FIGURE 7 – *FURIG* : Intensités moyennes en valeurs absolues selon le type catégoriel de sonde présenté aux modèles MINERVA2 et ATHENA

Résultats. Par souci de clarté — et bien que la moyenne de l'intensité des exemplaires d'une nouvelle catégorie soient toujours négative — nous présentons ces résultats dans la figure 7 en valeurs absolues afin de pouvoir comparer visuellement l'intensité de familiarité à l'intensité de surprise.

Pour MINERVA2, nous pouvons voir que l'intensité moyenne des prototypes (0.49(0.16)) est inférieure à l'intensité moyenne en valeur absolue des exemplaires de la nouvelle catégorie (-0.85(0.22)). Cette différence est significative : ($T(9998) = 92.239$; $p < 0.001$). Pour ATHENA, la même chose se retrouve avec une intensité moyenne des prototypes (0.62(0.04)) inférieure à l'intensité moyenne en valeur absolue des exemplaires de la nouvelle catégorie (-0.63(0.04)). Avec une différence significative : ($T(9998) = 13.537$; $p <$

0.001). La différence de résultats est, à nouveau, visiblement plus marquée pour MINERVA2 que pour ATHENA, ce qui est dû au fait que certaines intensités des exemplaires de la nouvelle catégorie dépassent largement -1 (tandis que l'intensité d'ATHENA est bornée). Cette différence ne remet pas en cause la pertinence du pattern de réponses, et la distinction catégorielle de Brouillet & Jimenez se retrouve dans les deux modèles : la nouveauté est plus surprenante que ce que la prortotypie est familière. MINERVA2 y parvient par abstraction d'exemplaires représentationnalistes, tandis qu'ATHENA parvient à généraliser l'emploi de processus non représentationnels pour obtenir les mêmes résultats. L'emploi de représentations n'est pas nécessaire pour la comparaison d'émergences catégorielles abstraites.

En ce qui concerne la comparaison entre le prototype et les exemplaires de la catégorie apprise (le paradigme DRM), pour MINERVA2 nous pouvons voir que l'intensité moyenne des exemplaires appris (0.51(0.24)) est significativement supérieure à l'intensité moyenne du prototype ($T(29998) = 5.963$; $p < 0.001$). Tandis que pour ATHENA, l'intensité moyenne des exemplaires appris (0.59(0.06)) est significativement inférieure à l'intensité moyenne du prototype ($T(29998) = 39.92$; $p < 0.001$).

Nous avons pu voir dans la simulation du paradigme DRM que — bien qu'un prototype des exemplaires rencontrés ne soit pas appris — la présentation d'une information prototypique est autant (si ce n'est plus) fluente que la présentation des exemplaires appris. C'est une particularité propre au fonctionnement de MINERVA2 qui se retrouve dans ATHENA, mais nous pouvons voir ici que MINERVA2 ne peut pas généraliser le paradigme DRM à des distinctions de catégorisation par valeurs flottantes, tandis que ATHENA en est capable. L'emploi de valeurs flottantes pour quantifier l'activité d'une composante (au lieu de se contenter d'activations ou d'inhibitions) est efficacement géré par notre modèle : une faible activation de composante est différente d'une forte activation de la même composante. Ce qui permet à notre modèle d'effectuer un appariement global de valeurs flottantes et de reproduire le pattern du paradigme DRM.

3.1.4 Memory-Conjunction Errors

Avec la simulation du paradigme DRM nous avons pu montrer que les capacités induites par l'ecphorie synergétique de MINERVA2 se retrouvaient dans le modèle ATHENA. Cependant, si ces capacités sont issues de l'utilisation de représentations en mémoire pour MINERVA2, ATHENA y parvient sans utiliser de représentations, mais il reste à s'assurer que l'utilisation d'une mémoire sans représentation parvient aussi à gérer l'abstraction inter-composantes. Pour ce faire, nous allons simuler les erreurs de type MCE (memory-conjunction errors).

Expérience originelle. Les erreurs dues aux conjonctions illusoires (MCE : Reinitz, Lammers & Cochran, 1992) sont typiques d'une diminution de l'activation intra-

composante au bénéfice de l'activation inter-composantes. Ce sont des erreurs assez courantes et qui se manifestent lorsqu'on croit se souvenir d'une situation pourtant nouvelle à cause d'une recombinaison d'éléments déjà connus sur la base de la mise en contexte de ces éléments. L'impression de déjà-vu en est un bon exemple. L'explication généralement admise est que ce phénomène serait lié à une augmentation inappropriée de la familiarité qui entraînerait un sentiment de familiarité pour l'événement pourtant nouveau (O'Connor & Moulin, 2008, 2010).

Dans cette expérience, les sujets sont soumis à une tâche d'apprentissage de 24 non-mots bisyllabiques composés de 48 syllabes uniques générées aléatoirement. Lors de la phase de reconnaissance, 18 non-mots sont présentés aux sujets : 6 proviennent de la liste d'apprentissage (mots anciens), 6 proviennent de la recombinaison de syllabes utilisées dans la liste d'apprentissage (mots nouveaux recombines), et 6 proviennent de la recombinaison d'une syllabe présente dans la liste d'apprentissage avec une nouvelle syllabe (mots nouveaux partiellement recombines). Les 38 syllabes utilisées lors de la phase de reconnaissance sont uniques.

Les résultats ont pu montrer que la perception de mots nouveaux recombines donnait lieu à d'avantage de fausses reconnaissances que la perception de mots nouveaux partiellement recombines. Ce résultat est interprété comme une erreur de recombinaison d'un mot pourtant nouveau. Cette erreur serait liée à une mauvaise recombinaison des traces, et donc à une abstraction inter-composantes illusoire. C'est pourquoi nous pensons que cette expérience est une bonne candidate pour mettre à l'épreuve MINERVA2 et ATHENA car ces deux modèles — s'appuyant sur le processus d'ecphorie synergétique — devraient pouvoir rendre compte des résultats expérimentaux.

Procédure de simulation. Cette expérience, à notre connaissance, n'a pas été simulée par MINERVA2. Pour pouvoir simuler cette expérience, 24 premières syllabes uniques et 24 secondes syllabes uniques sont générées à partir de vecteurs de 5 composantes chacune. Puis les non-mots de la liste d'apprentissage sont générés en concaténant les premières syllabes aux secondes pour un total de 24 vecteurs de 10 composantes. Pour la génération des 18 vecteurs de la tâche de reconnaissance, 6 proviennent de la liste d'apprentissage (mots anciens), 6 proviennent de la recombinaison de syllabes utilisées dans la liste d'apprentissage (les cinq premières composantes d'un vecteur sont concaténées aux cinq dernières d'un autre vecteur : ce sont les mots anciens recombines), et 6 proviennent de la recombinaison d'une syllabe présente dans la liste d'apprentissage avec une nouvelle syllabe différente des 48 générées en début d'expérience (mots nouveaux partiellement recombines). Au final, les 38 syllabes utilisées lors de la phase de reconnaissance sont différentes les unes des autres.

Pour analyser la réponse de reconnaissance des modèles, nous en récupérons la valeur d'intensité pour chaque présentation de sonde dans la phase de reconnaissance. Une valeur

positive est propre à la reconnaissance de la sonde (réponse « ancien »), tandis qu'une valeur négative nous informe sur la non-reconnaissance de la sonde (réponse « nouveau »). Les simulations sont reproduites 5000 fois pour chaque modèle, nous serons ainsi en mesure d'analyser le pourcentage de fausses reconnaissances pour les mots nouveaux selon leur caractéristique « recombinaire » (recombinés ou partiellement recombinaires), et ainsi comparer les résultats des deux modèles aux résultats expérimentaux.

Prédictions. Les mots anciens devraient être plus souvent reconnus que les mots nouveaux recombinaires, lesquels devraient également être plus souvent reconnus que les mots nouveaux partiellement recombinaires.

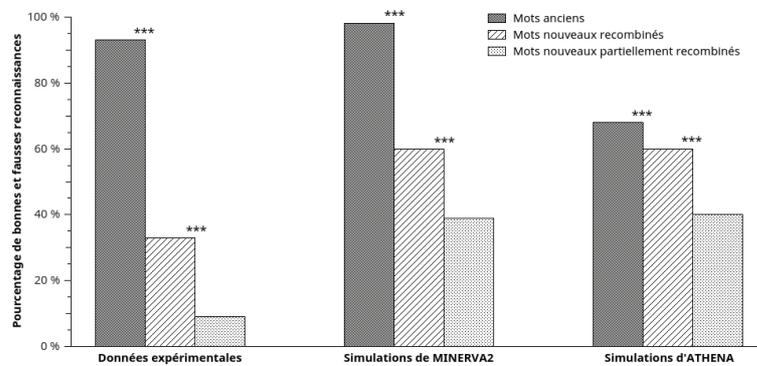


FIGURE 8 – MCE : Pourcentage de bonnes et fausses reconnaissances selon le type de mot pour les données expérimentales et les simulations des modèles MINERVA2 et ATHENA

Résultats. Bien que les résultats de la figure 8 semblent différents pour l'expérience originelle, MINERVA2 et ATHENA, nous pouvons remarquer que le pattern de réponse reste globalement le même. Pour MINERVA2, le pourcentage de reconnaissances correctes (97.99 %) des mots anciens est supérieur au pourcentage de fausses reconnaissances des mots nouveaux recombinaires (59.85 %) de manière significative ($\chi^2(1) = 2766.632; p < 0.001$). Et le pourcentage de fausses reconnaissances des mots nouveaux recombinaires est supérieur au pourcentage de fausses reconnaissances des mots nouveaux partiellement recombinaires (38.38 %) de manière significative ($\chi^2(1) = 24570.202; p < 0.001$). De même, pour ATHENA le pourcentage de reconnaissances correctes (67.54 %) des mots anciens est supérieur au pourcentage de fausses reconnaissances des mots nouveaux recombinaires (59.14 %) de manière significative ($\chi^2(1) = 455.805; p < 0.001$). Et le pourcentage de fausses reconnaissances des mots nouveaux recombinaires est supérieur au pourcentage de fausses reconnaissances des mots nouveaux partiellement recombinaires (40.55 %) de manière significative ($\chi^2(1) = 2073.549; p < 0.001$). Les erreurs recombinaires sont effectives pour les deux modèles et permettent de s'assurer que les recombinaisons illusoire sont modélisables par l'emploi de l'écphorie synergétique en inter-composantes.

En comparant les résultats de MINERVA2 à ceux de ATHENA, nous pouvons remar-

quer que si les mots nouveaux recombinaison ne sont pas différemment reconnus d'un modèle à l'autre ($\chi^2(1) = 3.138$; $p = 0.077$), le pourcentage de bonnes reconnaissances de mots anciens est plus élevé pour MINERVA2 que pour ATHENA ($\chi^2(1) = 9750.073$; $p > 0.001$), et le pourcentage de fausses reconnaissances est plus élevé pour ATHENA que pour MINERVA2 ($\chi^2(1) = 29.566$; $p < 0.001$). ATHENA est sujet à un pourcentage d'erreurs de reconnaissances plus élevé que MINERVA2.

En complément de la simulation du paradigme DRM, cette simulation nous montre que les capacités de MINERVA2 liées à l'emploi de l'ecphorie synergétique sont effectives pour ATHENA. Cependant, si MINERVA2 y parvient par la recombinaison concrète de contenus représentationnels, ATHENA y parvient par une contextualisation de covariances non-représentationnelle. Cette contextualisation complexifie grandement les interactions qui ont lieu entre les traces : le modèle ATHENA se construit par inférences de corrélations, ce qui — comparativement à MINERVA2 — permet une plus grande flexibilité mais provoque aussi une tendance plus importante à faire des erreurs d'attributions.

3.1.5 Conclusion

Dans cette partie nous avons pu mettre ATHENA à l'épreuve sur les tâches mnésiques propres au fonctionnement de MINERVA2 pour tester la capacité d'une mémoire non représentationnelle à rendre compte du processus d'ecphorie synergétique (processus d'appariement global).

La simulation du rappel indicé était nécessaire pour s'assurer que ATHENA parvient à reproduire les comportements fondamentaux attendus d'un modèle de mémoire, à savoir la récupération d'informations complémentaires à une situation incomplète, ainsi qu'une amélioration des capacités mnésiques par répétition. En premier lieu, lorsque les modèles doivent effectuer le rappel indicé d'une information à laquelle ils n'ont jamais été confrontés, MINERVA2 génère une réponse ayant les mêmes caractéristiques qu'une réponse aléatoire, tandis que ATHENA tente de construire une réponse en fonction de ce qu'il connaît (ce qui est forcément différent de la cible et provoque une réponse différente du hasard). Ensuite, l'amélioration des capacités mnésiques par répétition donne lieu à des formes de courbes différentes : MINERVA2 améliore sa mémoire de manière linéaire tandis que ATHENA l'améliore de manière logarithmique. Nous avons pu voir qu'en ce domaine, ATHENA parvient non seulement à reproduire les capacités de MINERVA2, mais aussi que sa courbe d'apprentissage est plus proche de celle d'un sujet humain.

La simulation de l'effet FURIG a permis de s'assurer que ATHENA se comporte conformément à une mémoire épisodique à système unique, et ce bien que ce modèle ne mémorise pas les situations sous la forme de traces épisodiques représentationnelles. En effet, il ne mémorise pas chaque exemplaire comme une nouvelle trace, mais modifie sa structure (sa manière de traiter l'information) de façon à inclure le traitement d'un

exemplaire dans l'ensemble de ses traces. Bien que le modèle ATHENA ne se souvienne pas de contenus représentationnels mais de processus réutilisables, il est tout de même un modèle de mémoire épisodique qui reproduit le même comportement que son équivalent représentationnaliste, et a fortiori le comportement humain.

La simulation de l'adaptation à la nouveauté issue de la modification de l'expérience de Whittlesea (1987) par Brouillet & Jimenez (2007) sur la prototypie, a permis de s'assurer d'une capacité propre à la mémoire humaine : la nouveauté est plus surprenante que ce qu'un prototype est connu. Simuler une telle expérience nous a demandé de sortir MINERVA2 de sa zone de confort par l'emploi de valeurs flottantes, mais celui-ci a obtenu des résultats tout de même pertinents. Ce modèle n'a pas pu généraliser le paradigme DRM à une abstraction plus fine de contenu, mais MINERVA2 n'était prévu que pour fonctionner dans ce contexte expérimental particulier. L'efficacité de ATHENA pour simuler cette expérience nous a non seulement confirmé l'idée que son adaptation à la nouveauté est cohérente avec les attentes expérimentales, mais aussi que l'emploi de valeurs flottantes est suffisamment bien géré par le modèle pour reproduire les effets du paradigme DRM avec un niveau plus fin d'abstraction catégoriel (i.e., une différenciation intra-composante au cœur d'une autre différenciation intra-composante).

La simulation des recombinaisons illusoirs est un complément à la simulation du paradigme DRM. Si l'association intra-composante dans l'abstraction mnésique a pu être simulée par ATHENA dans l'article du chapitre précédent, il était nécessaire de montrer que ATHENA parvient également à gérer l'association inter-composantes dans l'abstraction mnésique. Le fait que le modèle ATHENA en soit capable indique que l'écphorie synergétique n'a pas nécessairement besoin de traiter du contenu représentationnel pour y parvenir.

En définitive, ATHENA reproduit les mêmes capacités mnésiques que le modèle MINERVA2 sans avoir recours à l'emploi de représentations en mémoire. Certaines différences de comportements restent tout de même à noter, que ce soit au bénéfice d'ATHENA (courbe d'apprentissage logarithmique et catégorisation efficace de valeurs flottantes) ou au bénéfice de MINERVA2 (distinctions plus visibles et tendance moins marquée à faire des erreurs d'attributions).

3.2 Comportement temporel

Maintenant que nous avons pu vérifier que ATHENA conservait les mêmes caractéristiques mnésiques de base que MINERVA2 — son équivalent représentationnaliste — il est nécessaire de mettre ces modèles à l'épreuve sur leur différence comportementale fondamentale : l'aspect temporel. En effet, dans l'état de l'art nous avons mis en avant le fait qu'une mémoire énonciviste (i.e., sans représentation) devait acquérir un compor-

tement temporel cohérent. Afin de s'en assurer, quatre simulations expérimentales vont être présentées dans ce chapitre.

La première simulation (tâche d'interférence relative) a pour but de s'assurer que l'amorçage temporel présenté dans la simulation de l'expérience de Heurley et al. (2013) est contextuellement cohérent : un amorçage positif se produit dans une situation contextuellement congruente à la tâche, et un amorçage négatif se produit dans une situation contextuellement incongruente à la tâche. La deuxième simulation (effet de primauté/récence) a pour but d'analyser les relations temporelles d'un ensemble de sondes présentées de manière sérielle, et ce afin de s'assurer qu'un comportement temporel généralement considéré comme issu d'une modularité mnésique (i.e., mémoire à long terme vs mémoire à court terme) peut émerger d'une mémoire énacliviste à système unique. La troisième simulation (réverbération mnésique) a pour but d'analyser la persistance d'un traitement dans les traitements suivants de manière temporellement cohérente : cette simulation est caractéristique d'une interférence proactive et est en mesure d'expliquer l'effet de primauté. Enfin, la quatrième simulation (prédiction mnésique) a pour but d'analyser la cohérence temporelle de la modification d'anciens souvenirs par la création de nouveaux souvenirs : cette simulation est caractéristique d'une interférence rétroactive et est en mesure d'expliquer l'effet de récence.

3.2.1 Interférence relative

En montrant que le traitement d'une information pouvait modifier le traitement des informations suivantes, la simulation de l'expérience de Heurley et al. (2013) a pu montrer une certaine forme de temporalité dans le traitement mnésique du modèle ATHENA. Cependant, si le but de cette simulation était de montrer que ce modèle apprend la manière dont il traite l'information, elle ne montre qu'une dissociation entre condition congruente et condition incongruente (sans pouvoir définir s'il s'agissait de l'amorçage positif d'une situation congruente ou de l'amorçage négatif d'une situation incongruente). Il serait nécessaire de présenter au modèle une amorce neutre afin de pouvoir faire la distinction entre les deux types d'amorçage, et définir si l'effet de l'amorçage suit une logique temporelle cohérente en fonction de la pertinence de la situation.

Expérience originelle. L'emploi d'une amorce neutre pour quantifier les effets des amorces positives et négatives a été proposé par Neely (1977). Pour ce faire, l'auteur présente aux sujets une amorce de type catégorielle (le nom d'une catégorie) avant de leur demander une catégorisation mot/non-mot d'un item cible (le nom d'un exemplaire de catégorie). L'amorçage peut prendre trois formes : une amorce positive en présentant un exemplaire de la catégorie (e.g., OISEAUX-pinson), une amorce négative en présentant un exemplaire d'une autre catégorie (e.g., OISEAUX-porte), ou une amorce neutre en précédant l'exemplaire d'un non-mot (e.g., XXX-porte). Les résultats montrent qu'en

comparant les temps de réponse des situations congruentes et incongruentes avec ceux de la situation contrôle, un effet d'amorçage positif a lieu en cas de situation congruente, et un effet d'interférence négatif a lieu en cas de situation incongruente. Cette expérience montre qu'il existe un effet de cohérence contextuelle dans la temporalité du traitement.

Procédure de simulation. Pour simuler cette expérience nous utilisons une procédure similaire à celle mise au point pour simuler l'expérience de Heurley et al. (2013). Nous employons des vecteurs de 10 composantes dont les cinq premières permettent de présenter une catégorie et les 5 dernières permettent de présenter un exemplaire. Trois catégories différentes sont générées aléatoirement à partir de cinq valeurs entières (i.e., -1 et 1) en utilisant la méthode de différenciation présentée dans le chapitre précédent. Une différenciation maximum de -0,15 entre tous les vecteurs est utilisée de manière à obtenir des catégories les plus différentes possible sans ralentir outre mesure le calcul de la simulation. De la même manière, deux exemplaires différents de cinq composantes sont générés aléatoirement.

Afin de simuler la connaissance de corrélation entre les catégories et les exemplaires, une phase d'apprentissage présente aux modèles MINERVA2 et ATHENA deux sondes concaténant un exemplaire et une catégorie différente (e.g., OISEAUX-pinson et BATIMENT-porte). De cette manière, ces deux catégories sont corrélées à leurs exemplaires et la troisième catégorie inconnue des modèles représente le non-mot (amorce neutre).

Trois groupes sont utilisés pour la phase expérimentale : Le premier se voit présenter une amorce positive avant de traiter l'item correspondant (OISEAU-pinson), le deuxième se voit présenter une amorce négative avant de traiter l'item correspondant (BATIMENT-pinson), et le troisième se voit présenter une amorce neutre avant de traiter le même item (XXX-pinson). Nous ne pouvons pas demander aux modèles de catégoriser les items selon leur nature mot/non-mot, mais dans la mesure où cette tâche de catégorisation s'effectue en fonction de la familiarité de l'item, nous pouvons observer l'intensité du traitement effectué afin de quantifier la fluence de traitement (cf. Whittlesea, 1997). Les simulations sont reproduites 5000 fois afin de pouvoir effectuer un traitement statistique du comportement des modèles.

Comme pour la simulation de l'expérience de Heurley et al. (2013), la familiarité est inversement corrélée au temps de réponse. Nous prenons alors l'indicateur (1 - Intensité) afin de comparer les résultats des modèles aux données expérimentales dans la figure 9.

Prédictions. MINERVA2 ne fait pas de distinction temporelle entre ses traces et n'apprend pas la manière dont il traite l'information. Dans la mesure où la différence entre amorce positive, neutre ou négative dépend du lien existant entre l'amorce et la cible, les différents groupes devraient obtenir les mêmes résultats dans ce modèle.

L'inter-dépendance des traces du modèle ATHENA devrait le rendre sensible à la cohérence de la dynamique temporelle entre les situations rencontrées selon leurs liens de

corrélation. Dès lors, le traitement d'une situation devrait modifier le traitement de la situation suivante, et nous devrions obtenir des résultats cohérents avec le type d'amorce utilisé dans les trois groupes.

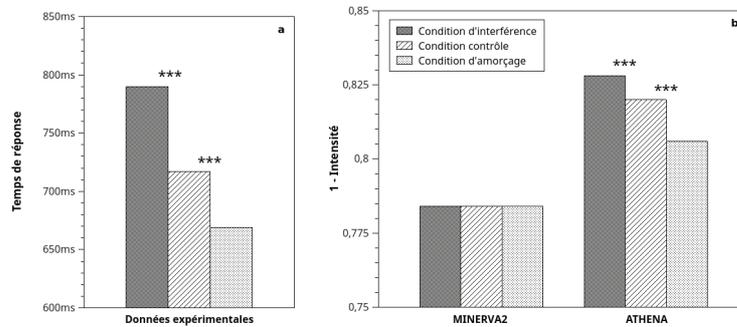


FIGURE 9 – *Interférences* : Temps de réponse selon le type d'amorçage pour les données expérimentales et les simulations des modèles MINERVA2 et ATHENA

Résultats. Nous pouvons remarquer dans la figure 9 que le modèle MINERVA2 produit les mêmes résultats pour les trois groupes, que ce soit au niveau de la distinction amorce_positive/amorce_neutre ($T(9998) = 0$), comme pour la distinction entre les conditions amorce_négative/amorce_neutre ($T(9998) = 0$). Les valeurs nulles des tests de Student sont issues d'une égalité absolue entre les différents groupes : le type d'amorce n'a strictement aucune influence sur le traitement de l'item cible.

ATHENA obtient par contre des résultats dont le pattern est conforme aux données expérimentales. La condition d'interférence (amorce négative) est significativement supérieure à la condition neutre ($T(9998) = 6.588$; $p < 0.001$), et la condition d'amorçage positive est significativement inférieure à la condition neutre ($T(9998) = 3.388$; $p < 0.001$).

Si MINERVA2 n'a pas de comportement temporel dans cette simulation, nous pouvons remarquer que ATHENA parvient à la fois à effectuer un amorçage positif en cas de situation congruente, ainsi qu'un amorçage négatif en cas de situation incongruente, reproduisant ainsi le pattern de résultats des données expérimentales.

La logique temporelle des interactions entre les traces du modèle ATHENA permet un traitement cohérent des situations en fonction du type d'amorce. Ce comportement est issu des caractéristiques fractales des traces puisqu'elles traitent l'information comme un processus réutilisable et non pas comme un contenu récupérable. Cependant, si ce comportement temporel cohérent entre deux situations consécutives nous permet de nous assurer qu'un traitement énonciviste non-représentationnel conserve une cohérence des covariances contextuelles, il nous reste à observer en détails les caractéristiques temporelles des interactions s'opérant entre toutes les situations rencontrées.

3.2.2 Effet de Primauté/Récence

Dans l'analyse du traitement temporel opéré par un système mnésique, il est une expérience typique que MINERVA2 devrait être incapable de simuler : l'effet de primauté-récence (*serial position effect*, voir Murdock, 1968). La capacité de rappel d'une information est dépendante de sa position dans une liste d'apprentissage. Les situations apprises par MINERVA2 étant indépendantes, ce modèle n'est pas capable d'opérer une opération de récupération mnésique différente selon la position temporelle de l'apprentissage de l'item. Théoriquement, la non-indépendance des traces développée dans le modèle Act-In devrait permettre d'obtenir des effets temporels cohérents. Dans la mesure où le modèle ATHENA reprend ce principe sous la forme d'une fractalité du contenu/processus mnésique, la non-indépendance de ses traces devrait voir son comportement marqué par une logique temporelle émergente et cohérente. Nous allons pouvoir mettre à l'épreuve les liens temporels qui se créent entre les traces de par leurs inter-dépendances mutuelles construites au cours du temps à travers la simulation de l'effet de primauté-récence.

Expérience originelle. L'expérience de Murdock (1968) est simple, efficace, et ne demande que peu de matériel. Une liste de mots est présentée aux sujets dans une phase d'apprentissage, puis un rappel libre leur est demandé. La probabilité de rappel d'un mot est analysé en fonction de sa position dans la liste. Les premiers mots sont mieux rappelés que ceux en milieu de liste (c'est l'effet de primauté), et les derniers mots sont aussi mieux rappelés que ceux en milieux de liste (c'est l'effet de récence). Cet effet est généralement considéré comme un argument en faveur d'une vision modulaire de la mémoire où la mémoire de travail contient des informations qui se dégradent et modifient la récupération d'informations en mémoire à long terme (Atkinson & Shiffrin, 1968). Cependant, il a aussi été postulé (McCrary & Hunter, 1953) que cette courbe typique peut être due à deux mécanismes d'interférence simultanés : l'interférence proactive (les anciens souvenirs perturbent l'apprentissage de nouveaux souvenirs) et l'interférence rétroactive (les nouveaux souvenirs perturbent la récupération des anciens). La théorie de la mémoire modulaire est généralement considérée comme plus à même d'expliquer les résultats expérimentaux, cependant Franklin & Mewhort (2015) ont pu montrer que l'effet de primauté-récence était possible dans une mémoire à système unique en cas de non indépendance des traces. De même, Damas, Mille & Versace (2002) ont pu mettre au point une amélioration de MINERVA2 (MINERVA 2 Étendu) capable de reproduire l'effet de primauté/récence grâce à une modulation temporelle de l'apprentissage. Cependant, ces modèles étant fondamentalement des modèles représentationnels, il serait intéressant de voir si cette non-indépendance de contenu mnésique est aussi applicable à une non-indépendance de processus mnésiques. Et si un double mécanisme d'interférence serait bien capable d'expliquer l'ensemble de ces résultats.

Procédure de simulation. Nous avons pris en exemple l'expérience de Murdock pour

imager l'effet de primauté-récence car il s'agit d'une des expériences les plus connues et les plus visuelles (avec la fameuse courbe en U caractéristique de cet effet), mais nous ne nous baserons pas sur un rappel libre pour simuler cet effet car MINERVA2 et ATHENA sont incapables d'effectuer un rappel libre. Ce n'est pas un problème car l'effet de primauté-récence se retrouve aussi dans une situation de reconnaissance (à condition que la présentation des items soit suffisamment longue, Paivio & Csapo, 1969), ainsi qu'en termes de temps de réponse (Morin, Derosa, & Stultz, 1967), ce que nous serons en mesure de simuler sur MINERVA2 et ATHENA en observant leur valeurs d'intensité. Pour s'assurer que la reconnaissance d'un item ne viendra pas perturber la reconnaissance des autres, les modèles seront définis en mode « test » lors de la phase de reconnaissance (aucune information de traitement ne sera apprise, et chaque item sera considéré comme le premier item présenté aux deux modèles).

La phase d'apprentissage consiste en la présentation successive de huit sondes différentes composées de huit composantes chacune. Chaque sonde est présentée deux fois d'affilé pour s'assurer que l'effet de primauté-récence ressorte suffisamment (cf Paivio & Caspo, 1969). La phase de reconnaissance consiste en la présentation de chacune des sondes d'apprentissage pour en observer leurs valeurs d'intensité. Chaque expérience est reproduite 5000 fois et les intensités sont normalisées par une transformation min-max en intra-sujet pour s'assurer que les résultats soient comparables d'un sujet à l'autre : dans la mesure où ce sont les différences d'intensité entre chaque sonde pour chaque participant qui est observé, il est nécessaire de les borner entre 0 et 1 pour pouvoir analyser les résultats de l'ensemble des simulations.

Deux groupes sont créés afin d'utiliser les deux types de catégorisation à notre disposition dans les deux modèles : intra-composante et inter-composantes. La différenciation intra-composante consiste en l'utilisation de l'ensemble des composantes vectorielles pour chacune des sondes : des valeurs aléatoires entières leur sont attribuées jusqu'à ce que toutes les sondes de la liste soient différentes (selon la méthode de différenciation précisée au chapitre précédent). La différenciation inter-composantes consiste quant à elle en l'utilisation de composantes différentes pour différencier les vecteurs : des composantes aléatoires sont utilisées pour définir les différents vecteurs. Par exemple, les sondes (1, 1, -1) et (-1, -1, -1) ont une différence intra-composante ; tandis que les vecteurs (1, 0, 0) et (0, 0, -1) ont une différence inter-composantes.

Prédictions. Du fait de l'indépendance de ses traces, MINERVA2 n'a aucune raison de voir son comportement changer en fonction de l'ordre de présentation des sondes. L'analyse statistique des intensités selon la position dans la liste devrait donner une courbe plate, sans lien de corrélation, quelle que soit la méthode de différenciation utilisée. En ce qui concerne ATHENA, la non-indépendance des traces devraient lui permettre d'obtenir des courbes en U (propres à l'effet de primauté-récence) quelle que soit la méthode de

différenciation utilisée.

Plus spécifiquement, la différenciation inter-composantes ayant comme caractéristique majeure de limiter les interactions entre les composantes (donc entre les traces), seules les spécificités temporelles devraient ressortir des résultats. Et dans la mesure où les nouveaux souvenirs se construisent à partir des anciens, une interférence proactive devrait dominer dans les résultats. L'effet de primauté devrait être plus marqué que l'effet de récence et s'accompagner d'un effet d'empreinte pour la première sonde (laquelle détermine fondamentalement comment le modèle va être capable de construire les souvenirs suivants).

Tandis que la différenciation intra-composante va permettre une interaction accrue entre les composantes (donc entre les traces), ayant comme conséquence d'augmenter l'effet d'interférence rétroactive des nouvelles traces sur les anciennes. Un effet de récence plus marqué que l'effet de primauté devrait se retrouver dans les résultats.

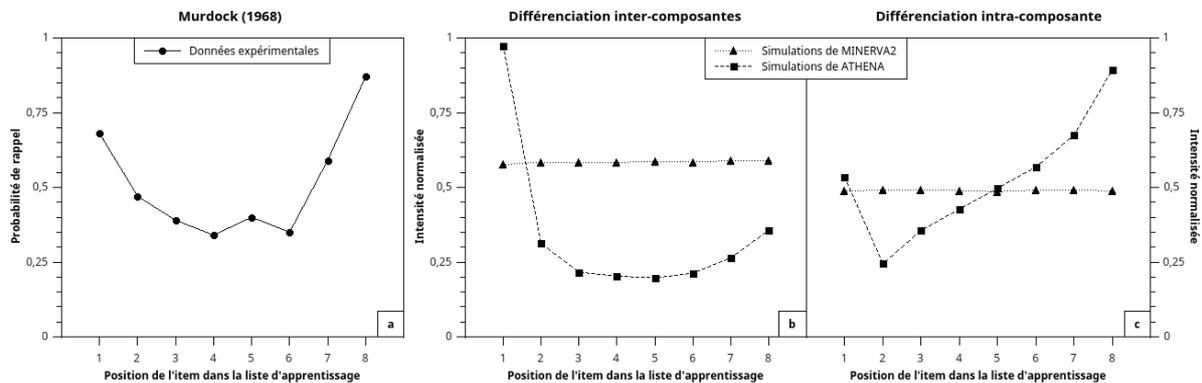


FIGURE 10 – *Serial Position Effect* : Probabilités de rappel et intensités normalisées selon la position sérielle de l'item pour les données expérimentales et les simulations des modèles MINERVA2 et ATHENA

Résultats. En premier lieu nous pouvons remarquer dans la figure 10 que MINERVA2 ne fait pas de distinction temporelle entre les sondes apprises. Ses résultats montrent une absence de corrélation entre la position d'une sonde apprise et l'intensité de sa reconnaissance, autant pour la différenciation inter-composantes (avec un test de Pearson à $r(40000) = 0.008$; $p > 0.05$) que pour la différenciation intra-composante (avec un test de Pearson à $r(40000) = 0.000$; $p > 0.05$). La différence apparente d'intensité moyenne entre les deux courbes s'explique par une interaction moindre entre les traces lorsqu'elles se différencient de manière inter-composantes (non pas lors de l'apprentissage, ce qui provoquerait un comportement temporel, mais lors de la reconnaissance par un processus d'appariement global). MINERVA2 est incapable de simuler la courbe en U caractéristique de l'effet de primauté-récence.

En ce qui concerne ATHENA, les résultats sont complètement différents. Nous pouvons voir sur le graphique l'émergence d'une courbe en U pour les deux types de différenciation (inter-composantes comme intra-composante), cependant les deux courbes ont des formes

différentes.

La courbe propre à la différenciation inter-composantes a une forme conforme à ce qu'on peut attendre d'un modèle ayant un comportement temporel, bien que l'effet de récence soit amoindri comparativement à l'effet de primauté (surtout si on compare cette courbe aux données expérimentales). De plus, un effet d'empreinte de la première trace est très marqué ($T(9998) = 159.304$; $p < 0.001$). L'effet de primauté est significativement plus marqué que l'effet de récence (avec un test de Pearson à $r(40000) = -0.33$; $p < 0.001$). Cette courbe est conforme à ce que donnerait une interférence proactive plus marquée qu'une interférence rétroactive.

La courbe propre à la différenciation intra-composante présente une forme moins conforme aux données expérimentales, mais présentant tout de même la forme d'un U (voire d'un V). L'effet d'empreinte est ici significativement présent ($T(9998) = 49.077$; $p < 0.001$), mais il reste moins important que l'effet de récence (avec un test de Pearson à $r(40000) = 0.435$; $p < 0.001$). Cette courbe est conforme à ce que donnerait une interférence rétroactive plus marquée qu'une interférence proactive.

L'implémentation de la non-indépendance des traces caractéristique à Act-In dans ATHENA lui permet d'obtenir une courbe de primauté-récence selon l'ordre d'apprentissage, et ce même si ce modèle n'apprend pas directement des exemplaires à modifier dans sa structure temporelle d'apprentissage. Cette caractéristique nous a poussé à toujours utiliser un ordre aléatoire de présentation des sondes d'apprentissage dans les simulations précédentes pour éviter les biais temporels. En définitive, une mémoire énacliviste sans représentation acquiert les mêmes caractéristiques temporelles qu'une mémoire modulaire.

3.2.3 Réverbération mnésique

La simulation reproduisant la courbe de primauté/récence nécessite d'approfondir l'origine de cet effet. Si l'interférence entre les traces est responsable de ce comportement, nous devrions observer une interférence proactive de la construction d'un souvenir dans la construction des souvenirs suivants (i.e., l'activité d'un traitement à un instant t se retrouve à l'instant $t + 1$, $t + 2$, etc. de manière décroissante). C'est un phénomène neurologique connu sous le nom de réverbération, lequel est généralement considéré comme une preuve biologique de l'existence d'une mémoire de travail (Fuster & Alexander, 1971; Miyashita & Chang, 1988; Goldman-Rakic, 1995; Miller, Erickson & Desimone, 1996). Cependant, Mongillo, Barak & Tsodyks (2008) ont pu montrer plus récemment que cette réverbération neuronale est effective sur l'ensemble du cerveau, la mémoire de travail ne serait-elle pas alors un comportement particulier de la cognition plutôt qu'un module indépendant ?

Pour pouvoir mettre ATHENA à l'épreuve sur une tâche de réverbération, et observer l'existence d'une interférence proactive dans son comportement mnésique, il va être

nécessaire de récupérer des informations non-demandées indépendamment des informations demandées. En effet, il va falloir observer la persistance d'une situation à l'intérieur du souvenir des situations suivantes : l'activité d'une composante au cœur même de la récupération d'une autre composante. Il n'existe à notre connaissance aucun protocole expérimental capable de quantifier simultanément la présence de plusieurs souvenirs dans la construction un autre souvenir, cette simulation ne repose alors pas sur la reproduction d'un protocole existant mais sur une adaptation du protocole de l'effet de primauté/récence. Nous n'aurons aucune donnée expérimentale à laquelle confronter les résultats des modèles, mais la modélisation nous offre des outils simples pour observer cette information sans équivoque.

Procédure de simulation. Cette simulation complète celle que nous avons développée pour reproduire la courbe de primauté/récence, nous nous baserons dès lors sur le protocole de simulation développé dans le chapitre 3.2.2. Mais dans la mesure où la différenciation intra-composante recouvre l'ensemble des composantes de l'écho (ne nous permettant pas de distinguer efficacement les souvenirs entre eux dans l'écho), nous utiliserons plutôt une différenciation inter-composantes afin d'être en mesure d'observer directement la présence d'un souvenir dans un autre. En effet, si les trois situations apprises se définissent sous la forme d'une liste de vecteurs $((1, 0, 0), (0, -1, 0), (0, 0, 1))$, la première composante sera propre à la génération du souvenir de la première sonde, la deuxième composante sera propre à la génération du souvenir de la deuxième sonde, et ainsi de suite. Si l'interférence proactive est effective, une nouvelle présentation de la dernière sonde $(0, 0, 1)$ provoquera la génération d'un écho comprenant à la fois le souvenir de la troisième sonde dans la troisième composante, mais aussi (de manière moins intense) les souvenirs des sondes précédentes dans les autres composantes.

Comme pour la simulation de l'effet de primauté/récence, la phase d'apprentissage consiste en la présentation successive d'une liste de sondes présentées deux fois d'affilé chacune, et la phase de reconnaissance consiste en la présentation de chacune des sondes d'apprentissage en mode « test » (pour que la reconnaissance d'une sonde n'influence pas la reconnaissance des autres sondes). Cependant, nous n'observons pas l'intensité de reconnaissance des sondes, mais le contenu de l'écho. À cette fin, les valeurs non nulles des composantes de chaque sonde d'apprentissage sont entières (soit 1, soit -1), ce qui nous permet de quantifier la qualité de la récupération du souvenir en fonction de la valeur absolue des composantes de l'écho. Par exemple, une composante ayant une valeur de 0.2 dans l'écho sera mieux reconnue que si elle était de 0.1. Nous serons ainsi en mesure de quantifier la présence du souvenir de la sonde présentée à l'instant t au cœur des souvenirs des sondes présentées aux instants $t + 1$, $t + 2$, etc.

Pour s'assurer que les particularités structurelles autres que la position temporelle des sondes n'influencent pas les résultats, le nombre de composantes non-nulles de chaque

sonde est générée de manière aléatoire (entre une et trois composantes). De même, la position de ces composantes dans la sonde est aléatoire. Les sondes utilisées sont au nombre de 6 et toutes contiennent entre 6 et 18 composantes (ce nombre est défini de manière aléatoire pour chaque nouvelle simulation et chaque sonde d'une même simulation contient le même nombre de composantes). Qu'une situation soit caractérisée par l'emploi d'une ou plusieurs composantes, une moyenne de leurs valeurs absolues dans l'écho est effectuée pour quantifier leur présence dans un autre souvenir. Les simulations sont reproduites 5000 fois pour les besoins de l'analyse statistique.

Prédictions. L'indépendance des traces représentationnelles de MINERVA2 devrait se traduire par une absence totale de souvenir autres que celui généré par la sonde de reconnaissance. Toutes les valeurs des composantes propres au souvenir de la sonde présentée à l'instant t devraient être nulles au cœur du souvenir des sondes présentées aux instants $t + 1$, $t + 2$, etc. En revanche, si l'effet de primauté dans le comportement temporel de ATHENA que nous avons observé dans la simulation portant sur l'effet de primauté-récurrence est bien dû à une interférence proactive, les souvenirs des sondes présentées aux temps $t + 1$, $t + 2$, etc. devrait comporter dans leurs écho le souvenir de la sonde présentée à l'instant t , et ce de manière plus intense selon leurs proximité temporelles (la présence du vecteur de l'instant t devrait être plus intense dans le souvenir du vecteur de l'instant $t + 1$ que dans celui de l'instant $t + 2$).

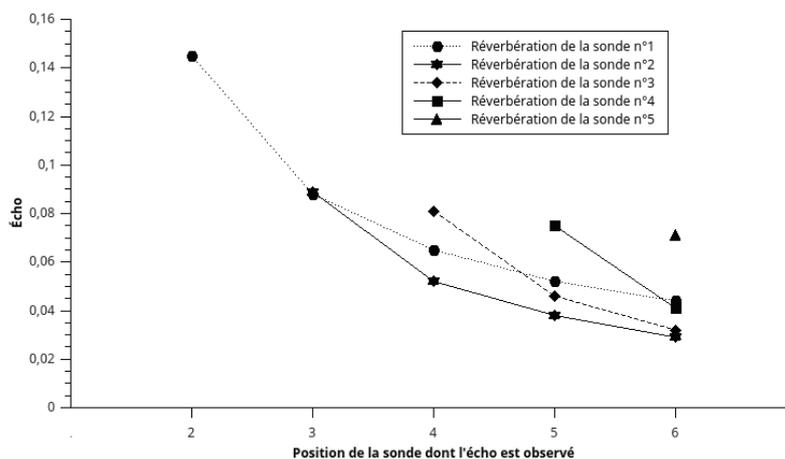


FIGURE 11 – *Interférence Proactive* : Valeurs moyennes des composantes d'autres souvenirs observées dans l'écho d'ATHENA

Résultats. Comme nous nous y attendions, la totalité des valeurs d'interférence proactives de MINERVA2 sont exactement à 0 (avec un écart-type de 0). Par souci de clarté, les valeurs nulles des résultats de ce modèle ne sont dès lors pas présentées dans la figure 11. L'indépendance des traces de ce modèle ne lui permet pas d'obtenir une quelconque interférence temporelle entre les souvenirs. Cependant, pour ATHENA nous pouvons remarquer dans la figure 11 que la génération d'un souvenir dans l'écho contient des valeurs

propres aux vecteurs d'autres situations que celle de rappel (bien qu'aucune corrélation autre que temporelle ne les relie). La présence d'un souvenir dans les suivants décroît avec la distance temporelle qui sépare les sondes propres à ces souvenirs, ce que nous pouvons vérifier avec un test de Pearson à $r(25000) = -0.43$; $p < 0.001$ pour la présence du souvenir de la première sonde dans les suivantes, un test de Pearson à $r(20000) = -0.40$; $p < 0.001$ pour la présence du souvenir de la deuxième sonde dans les suivantes, un test de Pearson à $r(15000) = -0.38$; $p < 0.001$ pour la présence du souvenir de la troisième sonde dans les suivantes, ainsi qu'un test de Pearson à $r(10000) = -0.3$; $p < 0.001$ pour la présence du souvenir de la quatrième sonde dans les suivantes (les cinquième et sixième sondes n'ont pas suffisamment de sondes suivantes pour effectuer un test de Pearson). Nous pouvons d'ailleurs remarquer que la réverbération de la première sonde recouvre certaines valeurs de réverbérations des autres sonde, ce qui est caractéristique de l'effet d'empreinte que nous avons pu observer dans les courbes de primauté/récence observées précédemment.

Cette simulation a pu montrer que l'activité mnésique engendrée par une situation fait l'objet d'une réverbération venant perturber la création des souvenirs suivants. Lors de la présentation d'une sonde au modèle ATHENA, la construction du souvenir ne dépend pas uniquement des inférences récursives des liens de covariances appris par le modèle (comme a pu le montrer la simulation de l'expérience de Brouillet et al., 2015a dans l'article présent dans le chapitre 2). La construction du souvenir dépend également des liens de corrélation temporels entre les situations apprises. Cette caractéristique est directement une conséquence de l'absence de représentation dans la mémoire : les souvenirs ne sont pas un contenu récupérable, mais une tentative de construire de l'information pertinente. Chaque trace va tenter de construire le souvenir selon sa propre histoire, et la mise en commun de l'ensemble des traces donne ce pattern cohérent. L'interférence proactive de réverbération n'est pas caractéristique de l'emploi d'une mémoire de travail, de même que l'effet de primauté n'est pas l'apanage d'un système mnésique modulaire. Ces phénomènes peuvent s'expliquer comme étant une même conséquence de l'inter-dépendance des traces d'une mémoire énonciviste.

3.2.4 Prédiction mnésique

Pour compléter la simulation précédente et mieux comprendre le pattern d'interférence obtenu dans la simulation de l'effet de primauté/récence, il est aussi nécessaire d'observer l'intervention d'une interférence rétroactive dans l'évolution temporelle de la mémoire. L'interférence rétroactive est une explication possible de la dégradation du souvenir même si l'interférence n'a aucun lien direct — autre que temporel — avec les items perturbés (Dewar, Cowan & Della Salla, 2007). L'interférence rétroactive serait en mesure d'expliquer l'effet de récence, là où l'interférence proactive peut expliquer l'effet de primauté (comme nous avons pu le voir en détail dans la simulation de réverbération mnésique).

Procédure de simulation. Pour pouvoir observer l'interférence rétroactive nous employons le même protocole de simulation que nous avons utilisé dans la simulation de l'interférence proactive. À la différence près que nous n'observons plus la présence d'un souvenir dans les souvenirs suivants mais la présence d'un souvenir dans les souvenirs précédents. Lors de la phase d'apprentissage, une liste de sondes différenciées de manière inter-composantes sont présentées aux deux modèles (chaque sonde est présentée deux fois d'affilée avant de passer à la suivante). Lors de la phase de reconnaissance, chaque sonde d'apprentissage est présentée aux modèles en mode test (pour s'assurer que la récupération d'un souvenir ne modifie pas la récupération des autres souvenirs). Nous observons alors la présence du souvenir de l'instant t dans l'écho des souvenirs des instants $t - 1$, $t - 2$, etc. de manière à quantifier la modification des anciens souvenirs par les nouvelles situations rencontrées. Les simulations sont reproduites 5000 fois pour pouvoir en analyser les particularités statistiques.

Prédictions. L'indépendance des traces représentationnelles de MINERVA2 ne lui permettant pas d'acquérir un comportement temporel, aucun lien ne sera construit entre les souvenirs, et la quantification d'un souvenir dans les souvenirs précédents devrait toujours être nulle : ce modèle ne devrait pas faire l'objet d'une interférence rétroactive. En ce qui concerne ATHENA, l'inter-dépendance de l'ensemble de ses traces permet aux souvenirs d'être encodés conjointement et la création d'un souvenir devrait modifier les souvenirs précédents. La conjonction de l'ensemble des interférences produites de manière sérielle devrait faire apparaître une cohérence temporelle de cet effet d'interférence rétroactive. Un souvenir sera présent dans l'ensemble des souvenirs précédents avec une intensité proportionnelle à la proximité de cette situation aux autres situations (le souvenir de la sonde présentée à l'instant t sera plus présent dans le souvenir de la sonde présentée à l'instant $t - 1$ que $t - 2$, etc.)

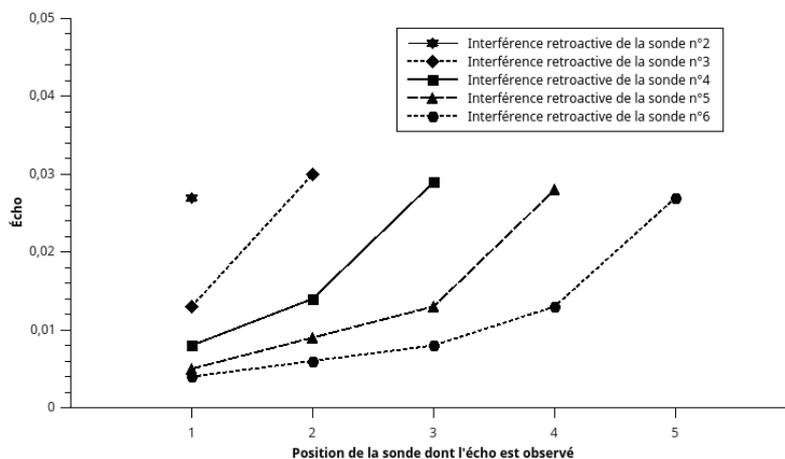


FIGURE 12 – *Interférence rétroactive* : Valeurs moyennes des composantes d'autres souvenirs observées dans l'écho d'ATHENA

Résultats. Comme nous nous y attendions, La totalité des valeurs d'interférence rétroactives de MINERVA2 sont à 0 (avec un écart-type de 0). Par souci de clarté les résultats de ce modèle ne sont pas présentés dans la figure 12. L'indépendance des traces de ce modèle ne lui permet pas d'obtenir une quelconque interférence temporelle entre les souvenirs. Cependant, pour ATHENA nous pouvons remarquer dans la figure 12 que la génération d'un souvenir dans l'écho ne contient pas uniquement les valeurs associées au souvenir rappelé, mais aussi des valeurs propres aux vecteurs d'autres situations (bien qu'aucune corrélation autre que temporelle ne les relie). La présence d'un souvenir dans les souvenirs précédents décroît avec la distance temporelle qui sépare les sondes propres à ces souvenirs, ce que nous pouvons vérifier avec un test de Pearson à $r(10000) = 0.323$; $p < 0.001$ pour la présence du souvenir de la troisième sonde dans les précédentes, un test de Pearson à $r(15000) = 0.378$; $p < 0.001$ pour la présence du souvenir de la quatrième sonde dans les précédentes, un test de Pearson à $r(20000) = 0.397$; $p < 0.001$ pour la présence du souvenir de la cinquième sonde dans les précédentes, ainsi qu'un test de Pearson à $r(25000) = 0.409$; $p < 0.001$ pour la présence du souvenir de la sixième sonde dans les précédentes (les première et deuxième sondes n'ont pas suffisamment de sondes précédentes pour pouvoir effectuer un test de Pearson).

Cette simulation a pu montrer que l'activité mnésique engendrée par une situation fait l'objet d'une interférence rétroactive venant perturber la récupération des souvenirs précédents. De ce fait, la re-présentation d'un souvenir entraînera automatiquement la re-présentation des souvenirs qui le suivent temporellement de manière cohérente. Par exemple, lorsqu'on rappelle un numéro de téléphone, le rappel des premiers chiffres entraîne le rappel des suivants. Il peut même être impossible de rappeler les derniers chiffres sans avoir au préalable rappelé les précédents. C'est en ce sens que l'interférence rétroactive peut être considérée comme une prédiction temporelle. Lors de la présentation d'une nouvelle sonde au modèle ATHENA, les souvenirs précédemment construits seront modifiés conformément à cette nouvelle situation et proportionnellement aux liens de corrélation temporels entre les situation apprises. Conjointement à l'effet de réverbération (interférence proactive), l'effet de réverbération inverse (interférence rétroactive) provoque une double interférence temporelle capable d'expliquer les spécificités de la courbe en U obtenue lors de la simulation de l'effet de primauté/récence.

3.2.5 Conclusion

La simulation de l'expérience de Heurley et al. (2013) sur l'amorçage laissait entrevoir un comportement temporel du modèle ATHENA dans le traitement de l'information. Dans ce chapitre nous avons pu tester plus en profondeur ses caractéristiques temporelles. MINERVA2 n'a pu reproduire aucun de ces comportements temporels.

La simulation de l'interférence relative a permis de nous assurer que ce comportement

temporel était cohérent avec le contexte de l'information. La dissociation entre traitement congruent ou incongruent à la situation n'est pas simplement le fait d'un amorçage seul, mais spécifiquement d'un amorçage positif ou négatif relativement à une situation neutre. Le lien temporel d'une situation à la suivante s'effectue de manière cohérente selon le type de situation rencontrée.

La simulation de l'effet de primauté-récence a permis de nous assurer que la cohérence temporelle ne s'effectuait pas simplement d'une situation à l'autre mais conjointement à l'ensemble des situations rencontrées et de leurs liens temporels relatifs. Si un effet de primauté/récence a déjà pu être modélisé dans une mémoire à système unique, cette simulation démontre que cet effet n'est pas nécessairement issu d'une double interférence des contenus mnésiques représentationnels qui se recouvreraient : l'énactivisme étant fondamentalement une théorie de la mémoire non-représentationnelle, le fait que le modèle ATHENA soit capable de reproduire ces résultats démontre que les représentations ne sont pas une nécessité dans le traitement temporel des situations rencontrées.

Pour mieux comprendre l'effet de primauté, la simulation de réverbération mnésique a permis de s'assurer qu'un effet d'interférence proactive avait bien lieu dans le fonctionnement du modèle ATHENA. De même, la simulation de prédiction mnésique a permis de s'assurer qu'un effet d'interférence rétroactive avait aussi lieu dans le fonctionnement du modèle ATHENA. Ces deux effets combinés permettent d'expliquer la courbe en U caractéristique de l'effet de primauté/récence, bien qu'aucune relation de corrélation autre que temporelle ne caractérise le lien présent entre les traces.

Afin de modéliser une mémoire énactiviste, nous avons utilisé une caractéristique dominante du modèle Act-In — à savoir une inter-dépendance totale entre les traces. Chacune des traces du modèle ATHENA est construite par l'interaction combinée des anciennes traces, et chaque trace utilisée dans l'interprétation de nouvelles situations est modifiée en fonction de la manière dont elle a su traiter l'information. Ces caractéristiques permettent au modèle ATHENA de développer un comportement temporel émergent. Ce qui montre que l'emploi de représentations en mémoire n'est pas nécessaire pour obtenir un comportement temporel cohérent ; et si l'on compare ce modèle à son équivalent représentationnaliste, nous pouvons même affirmer que la mémorisation de processus réutilisables en situation est plus efficace que la mémorisation de contenus récupérables.

Discussion

4.1 Énactivisme radical

Même si le thème est bien trop souvent ignoré dans la recherche en Psychologie — et plus généralement dans les sciences cognitives — la position prise sur la notion de représentation est fondamentale dans la manière de concevoir la cognition, et il n'existe pas de position neutre. La notion de représentation est originellement issue de l'héritage cognitiviste et plus particulièrement du dualisme cartésien. Cependant, si le dualisme cartésien est encore très peu défendu de manière ouverte, l'emploi de représentations n'est que rarement remis en question. La notion de représentation est pourtant porteuse de limites épistémologiques lourdes de conséquences (elles limitent l'emploi d'une structure cognitive capable d'évoluer, et il est nécessaire de rajouter au système cognitif un phénomène de transduction récuratif arbitraire). Si le connexionnisme a pu résoudre une partie des problèmes cognitivistes (la sérialité des processus et la localisation des symboles), il a été historiquement incapable de résoudre le problème des représentations. La remise en question totale d'un paradigme scientifique n'est jamais chose facile, et de nouvelles théories ne peuvent remarquablement voir le jour que si elles s'inscrivent dans le cadre des théories existantes. Le connexionnisme n'est au final qu'une mise à jour du cognitivisme alors que ses outils dynamiques (boucles de rétroactions, traitements en parallèle, etc.) et certains apports théoriques particulièrement forts (comme la notion d'émergence) auraient pu provoquer une véritable révolution scientifique si le connexionnisme ne s'inscrivait pas dans une logique représentationnaliste. L'emploi de représentations bride les apports du connexionnisme qui se retrouve au final à justifier la sérialité input/traitement/output par l'emploi de réseaux en couches de neurones, à avoir recours à un sub-symbolisme porteur des mêmes limites que l'emploi de symboles localisés, ou à oblitérer complètement les questions — pourtant essentielles — de conscience, d'ipséité, de perception temporelle, voire même de « soi ».

Si l'énactivisme s'est développé en réaction aux limites des représentations, la notion d'énaction n'échappe pas à ce principe. Différentes écoles ont émergé de l'apport d'une cognition incarnée et située, mais toutes ne respectent pas le rejet de la notion de représentation pourtant à son origine. L'énaction est souvent utilisée comme tremplin justifiant la dissociation cognitiviste faite entre les processus de bas niveau et de haut

niveau : tout ce qui a trait au traitement bas-niveau peut s'expliquer en termes d'énaction, mais tout ce qui a trait au haut-niveau s'expliquerait tout de même en termes de représentations. C'est une position que Hutto & Myin (2012) ont qualifiée d'énactivisme conservateur et qui perd toute cohérence avec le paradigme de Varela (1993, 1996) au profit d'une cohérence avec les théories cognitivistes. Comme le connexionnisme, l'énactivisme conservateur est une mise à jour du cognitivisme (dont l'aspect modulaire serait virtuellement capable d'intégrer n'importe quelle théorie). Pourtant, l'énaction est une conséquence du rejet du cognitivisme par un nombre croissant de chercheurs en sciences cognitives. Ces deux paradigmes sont incompatibles (Menary, 2010). Voilà pourquoi nous n'employons pas le terme énéaction mais énéactivisme dans ce travail : l'énéactivisme est un paradigme qui doit nécessairement se passer de l'emploi de représentations (Hutto & Myin, 2012, le nomment énéactivisme radical). Non seulement il manque à l'énéactivisme une modélisation mathématique de ses théories pour que ses apports puissent se diffuser dans la communauté scientifique, mais il manque surtout une modélisation de la cognition sans représentation (bien que techniquement un modèle énéactiviste soit par nécessité un modèle sans représentation). Ce que nous avons humblement tenté de réaliser dans ce travail.

L'énéactivisme radical (i.e., sans représentation — donc l'énaction originelle de Varela) est à même de résoudre le problème des représentations à condition de s'inscrire dans une logique plus phénoménologique de la cognition. Dans la mesure où la cognition ne se construit pas en fonction de la signification pré-existante du monde, mais en fonction des interactions qu'elle est capable d'avoir avec celui-ci, la distinction entre le monde et le sujet est superflue. La réunion monde-sujet permet de réduire à néant le problème de transduction à condition de se séparer d'un postulat très ancré dans notre tradition culturelle : l'unité ponctuelle et temporelle du « soi ». En effet, l'unité ponctuelle du « soi » est la croyance généralement admise qu'un individu possède une nature qui lui est propre, qui le caractérise et l'essentialise en tant qu'individu (dans le sens où l'individu est une individuation d'un ensemble de comportements) et qui le sépare du reste du monde (« je » suis ce que le reste n'est pas). L'unité temporelle du « soi » est la croyance généralement admise que l'unité ponctuelle du « soi » perdure au cours du temps (« je » suis aujourd'hui la même personne que hier, et serai aussi cette même personne demain). Si l'unité ponctuelle et temporelle du « soi » s'inscrit parfaitement dans une logique religieuse (qui essentialise le « soi » grâce à la notion d'esprit), il peut paraître curieux de la voir perdurer au sein des sciences cognitives alors que son existence n'a qu'une utilité descriptive. En effet, la persistance physique (autopoïese) ou psychologique (ipséité) n'est qu'une manière de décrire un système (voire pour un système de se décrire lui-même), et n'a un intérêt que pour l'interaction sociale (si l'on peut considérer que s'interpréter soi-même est une forme d'interaction sociale, puisqu'elle se fait en fonction d'autrui) : l'unité temporelle du « soi » est un postulat pratique mais n'a rien de scientifique (même si elle nous permet d'éviter

de nombreuses dissonances cognitives). De même, un système cognitif étant toujours en devenir, et n'étant qu'en recherche de stabilisation, l'unité ponctuelle du « soi » est aussi un postulat pratique : sa construction est très utile à étudier (pour déterminer comment le système se perçoit) mais sa *nature* est scientifiquement superflue. Une fois que l'on a retiré toute conception essentialiste à la nature du « soi », il devient plus facile de l'étudier sous la forme d'un comportement de « soi » : l'individu est l'interaction qui a lieu avec l'environnement (et non pas celui qui est en interaction avec son environnement). Il est important de ramener la notion d'autopoïèse — généralement pensée en terme d'une cellule unique — à la dimension macroscopique du sujet humain : de cette manière nous pouvons remarquer que si la distinction entre le sujet et le monde est erronée, la notion même de représentation est superflue puisqu'il n'est pas nécessaire de faire le lien entre l'individu et son environnement. De même, nous pouvons mieux comprendre en quoi toute forme de cognition est fondamentalement une activité. À partir de là, si l'individu est un comportement et la cognition une activité, la seule forme d'information mémorisable et utilisable par un système énactiviste sera l'évaluation d'une activité. Mais dans la mesure où une activité n'a aucune signification prise isolément, c'est par la confrontation de toutes les activités mesurables que la cognition parviendra à se construire : la covariance sensorimotrice. Au final, devoir faire appel à des représentations pour pouvoir utiliser un système cognitiviste est scientifiquement plus contraignant que se passer de représentations. La notion de « soi » n'est pas justifiable, et si l'on peut montrer que l'approche énactiviste est à même de reproduire les mêmes comportements que ceux prédits par le paradigme cognitiviste, le rasoir d'Ockham devrait reléguer au placard les théories cognitivistes.

Bien que ce travail ne porte pas sur l'activité de « soi », l'énactivisme est fondamentalement une science de l'expérience phénoménologique. Dès lors, ce paradigme doit être capable de prendre en compte le fait que la distinction entre le monde et le sujet n'est pas une donnée pour le système cognitif. Non seulement il doit la construire par son expérience « personnelle » de ses interactions sensorimotrices, mais cette idée pose les bases du principe selon lequel l'utilisation de la covariance n'est pas uniquement nécessaire pour l'interaction entre le monde et sujet, mais surtout entre les différentes expériences de l'interaction entre le monde et le sujet. Par exemple, l'activité simultanée d'une perception auditive et d'une perception visuelle n'a aucune raison d'être fondamentalement de nature différente : c'est la covariance entre les processus qui détermine les liens de corrélations (associatives ou dissociatives). Les processus cognitifs ne sont pas plus distincts entre eux que ce que le monde est différent du sujet. L'application de l'énactivisme à une vision phénoménologique du « soi » nous pousse fondamentalement à considérer que les processus cognitifs interagissent comme un couplage sensorimoteur (de même que l'interaction entre ces interactions cognitives, la récursivité permet une complexification du traitement cognitif sans avoir besoin de dissocier processus de haut et bas niveau). Tout comme l'absence de représentation en mémoire ou la réunion processus/contenu,

l'invariance d'échelle est une caractéristique de l'énaïvisme (le modèle Act-In, Versace et al., 2014, est alors fondamentalement un modèle énaïviste). La fractalité de la cognition développée dans ce travail peut sembler être un ajout théorique à l'énaïvisme, mais elle n'est en réalité que le prolongement de son application stricte.

4.2 Fondamentaux d'une mémoire énaïviste

Se passer de l'essentialisation d'un contenu cognitif improbable nous a permis de mettre en exergue quatre fondamentaux d'une mémoire énaïviste.

En premier lieu, la conception d'une mémoire sans représentation n'a aucune raison d'être modulaire. L'emploi d'un système mnésique unique et épisodique est la manière la plus pertinente de traiter des activités mnésiques non-essentialisées. Nous nous sommes dès lors basés sur MINERVA2 pour modéliser l'éphorie synergétique des traces mnésiques au contexte situationnel, à la différence près que ATHENA ne traite plus du contenu mnésique mais du traitement mnésique (l'information n'est pas récupérée en mémoire, elle est construite).

Le deuxième point fondamental provient directement de l'absence de représentation : la cognition ne se souvient pas de contenus récupérables mais de processus réutilisables. Baser l'ensemble de l'information mnésique sur la mémoire des covariances permet de se passer de contenu essentialisé (ainsi que de pattern représentationnel propre au connexionnisme). De cette manière, le problème de transduction récursive n'a pas lieu d'être, puisque l'information ne circule pas entre des modules arbitrairement distincts.

Le troisième point fondamental concerne l'utilisation que l'on peut faire de telles covariances apprises : la contextualisation. Puisqu'il n'y a plus d'information objectivement pertinente à récupérer, la construction d'un souvenir s'opère en adaptant la mémoire à la situation présente (ce qui résout le problème du sub-symbolisme connexionniste), ce qui est conforme à l'éphorie synergétique, mais aussi aux caractéristiques d'un système autopoïétique. En ce sens, l'émergence du souvenir n'est pas le but de la mémoire mais un effet secondaire de celle-ci. Si l'on considère la cognition de manière évolutionniste comme un système autopoïétique, sa seule fonction est de perdurer, de conserver une cohérence interne en évitant toute forme de dissonance cognitive. Dès lors, puisque le système cognitif subit constamment des perturbations provenant de l'environnement et de lui-même (pour un système non-isolé, qu'une attente perceptive perturbe un traitement visuel revient au même que si un traitement auditif perturbe ce traitement visuel) la réaction sensorimotrice du système est une forme de résonance dont la seule utilité est de rétablir un état d'équilibre suite à une perturbation. Cependant, dans la mesure où les perturbations sont constantes et sont en rétroaction perpétuelle, le système n'atteint jamais un état d'équilibre. Cette particularité a été étudiée par Sartre (1936) qui considérait que l'état actuel du système (en-soi) et l'état d'équilibre à atteindre (pour-soi) est un

jeu de miroirs se reflétant perpétuellement (jusqu'à la mort du sujet où techniquement les deux se rejoignent). Cependant, l'en-soi n'est pas un donné, il est inféré et est donc une interprétation qui ne peut se faire que par comparaison avec le pour-soi (on se définit en fonction de ce qu'on aimerait être) ; de même, le pour-soi n'est pas un donné et s'infère en fonction de l'en-soi (on détermine nos choix en fonction de l'état actuel du système). Ce double jeu en-soi/pour-soi où l'identité est constamment en devenir permet de se passer d'une hypothétique nature de « soi », et par extension la contextualisation de covariances sensorimotrices n'est pas incompatible avec la notion de construction de « soi » ni l'émergence d'une certaine forme d'intentionnalité (sur laquelle nous reviendrons). Mais dans la mesure où la construction de « soi » ne s'effectue pas de manière unitaire (ni ponctuellement ni temporellement), il est nécessaire d'appliquer à cette contextualisation une invariance d'échelle dans son fonctionnement.

Nous en sommes ainsi venus au quatrième point fondamental d'une mémoire énaïviste, qui, bien qu'elle en soit une application directe est encore peu étudiée : l'invariance d'échelle. Il est un modèle qui, s'il ne s'en revendique pas, est pourtant à ce jour le seul à pouvoir rendre compte d'une mémoire énaïviste : le modèle Act-In (Versace et al., 2014). En effet, ce modèle porte les germes d'une mémoire non-représentationnelle (non-indépendance des traces, invariance d'échelle et construction mnésique par le sensorimoteur). Dans une mémoire à traces multiples où chaque trace (ainsi que chaque composante de la trace) est un processus, l'absence de représentation n'est efficace que si les processus internes à la mémoire se comportent comme la mémoire dans sa globalité. Chaque processus tente de reconstruire le souvenir, et c'est la synergie de l'ensemble des tentatives de reconstruction qui permet l'émergence du souvenir. Dans ce cas, toutes les traces se modifient en fonction de la manière dont elles traitent l'information, et comme les traces se modifient toutes en fonction de covariances corrélées à la même situation (les traces se modifient en parallèle en fonction de leur manière de construire la nouvelle trace), il en résulte une interdépendance des traces capable de se souvenir des covariances contextualisée des processus mnésiques de manière récursive. Ainsi, la manière de traiter une information est construite, ce qui résout le problème de la structure pré-donnée des systèmes représentationnalistes. Cette particularité est la clé de voûte permettant à une mémoire énaïviste de ne pas se retrouver limitée à l'étude des interactions « purement » sensorimotrices de bas niveau, et de construire des corrélations d'abstraction de manière récursive au cours du temps.

Dans la mesure où la constance temporelle n'est pas pré-donnée au système, il est une caractéristique qu'un modèle énaïviste se doit de réaliser : un comportement temporel émergent. En effet, si la flèche du temps n'est pas ce sur quoi se base la mémoire pour obtenir un comportement temporel cohérent, mais une interprétation de la logique temporelle (l'ipséité), un système énaïviste doit a minima être capable de produire un comportement temporellement cohérent (sur lequel construire une ipséité individuelle).

L'inter-dépendance des traces permet une forme de consistance propre à l'interprétation d'une unité ponctuelle, mais leur évolution conjointe doit pouvoir rendre compte de l'interprétation d'une unité temporelle (à partir de ce que les traces étaient vers ce qu'elles seront). Afin de s'assurer que les théories énonctivistes sur lesquelles nous avons bâti ATHENA sont cohérentes, ce modèle ne doit pas seulement être capable de reproduire le fonctionnement d'une mémoire représentationnaliste, mais aussi de faire émerger une sérialité mnésique temporellement cohérente (bien qu'aucune forme de temporalité absolue ne soit apprise par le modèle).

Le développement probabiliste du modèle MINERVA2 nous a permis de cibler les modifications à apporter à l'utilisation représentationnaliste de l'ecphorie synergétique, et ce afin de mettre au point une formalisation mathématique de l'énonctivisme. Les quatre points fondamentaux d'une mémoire sans représentation sont formalisés tels que l'équation (6) respecte la confrontation de l'ensemble des traces avec la situation présente (ecphorie synergétique), l'équation (18) modérée par l'équation (16) s'assure que les nouvelles traces n'apprennent pas le contenu de la sonde ou de l'écho mais la manière dont l'information est traitée (covariance sensorimotrice), l'équation (8) permet quant à elle d'inférer le comportement d'une trace en fonction du comportement des autres (une contextualisation des covariances), enfin, l'équation (6) est porteuse de la récursion fractale nécessaire à l'invariance d'échelle de ces équations et à une réunification processus-contenu (une meta-contextualisation de covariances fractales). Afin de simplifier la modélisation, les équations (19), (20), (21) et (22) permettent une implémentation informatique claire. Le code source du modèle ATHENA est sous licence CECILL-B, ce qui le rend librement accessible, réutilisable et modifiable afin de faciliter sa diffusion.

4.3 Mise à l'épreuve d'une mémoire énonctiviste

Trois séries de simulations ont été utilisées dans ce travail pour mettre le modèle ATHENA à l'épreuve en le confrontant à son équivalent représentationnaliste MINERVA2. La première série de simulations porte sur l'émergence, l'apprentissage et la manipulation d'abstractions, lesquelles sont encore considérées comme un argument en faveur de la nécessité d'employer des représentations. La seconde série de simulations a permis de s'assurer que les apports du modèle MINERVA2 étaient toujours présents même sans employer de représentations. Quant à la troisième série de simulations, elle nous a permis de développer le comportement temporel du modèle ATHENA afin de vérifier que l'inter-dépendance des traces était bien effective de manière cohérente.

La première série de simulations portant sur l'abstraction décline en trois répliques d'expériences l'émergence d'une abstraction (le paradigme DRM), l'apprentissage d'une

abstraction émergente (l'amorçage perceptif), ainsi que la manipulation et la confrontation d'abstractions émergentes apprises (l'affordance). La simulation du paradigme DRM donne des résultats similaires pour MINERVA2 et ATHENA, à la différence près que MINERVA2 parvient à l'émergence d'une abstraction à partir de contenu mnésique représentationnel tandis que ATHENA y parvient par une abstraction de processus de covariances sensorimotrices. La simulation de l'amorçage perceptif n'était possible que si les modèles étaient capables d'apprendre une abstraction émergente (i.e., la manière dont ils traitent l'information, et non pas seulement l'information en soi). Seul ATHENA parvient à modifier son fonctionnement en fonction de son activité propre. La simulation d'affordance mettait à la fois en jeu l'émergence d'abstractions et la confrontation de ces processus (comme s'ils étaient de simples contenus mnésiques). À nouveau, seul ATHENA est capable de manipuler ses processus en cours de manière cohérente. Mémoriser la covariance sensorimotrice plutôt que des représentations mnésiques permet à ATHENA d'être une mémoire de processus. Ces processus sont déjà une activité et n'ont pas besoin d'être mis en relation de manière artificielle puisqu'ils interagissent selon leur propre covariance de fonctionnements. Nous avons pu voir que si MINERVA2 était théoriquement capable de reproduire ces simulations avec une intervention extérieure du chercheur (à définir selon le type de tâche demandée), ATHENA en est capable de manière autonome. C'est en ce sens qu'une mémoire énacliviste est plus efficace qu'une mémoire représentationnaliste dans l'abstraction de concepts.

La simulation d'affordance a la particularité de mettre en avant une capacité d'ATHENA qui n'est possible que dans une vision énacliviste de la mémoire : l'émergence en cascade. En effet, le fait de concevoir le contenu de la mémoire comme des symboles récupérables nécessite un processus de traitement de l'information distinct de l'information à traiter. Dès lors, une émergence ne peut produire d'autres émergences que si la consigne est clairement donnée au modèle représentationnaliste (e.g., la présentation successive de l'écho comme nouvelle sonde plusieurs fois d'affilée pour MINERVA2). Tandis que pour un modèle énacliviste, si un processus en active un autre, ce nouveau processus activé propagera automatiquement son activation à d'autres processus (et ainsi de suite). Cette propagation d'activation s'effectuera de manière cohérente à la situation originelle ayant donné lieu à ce comportement (un nouveau processus s'activera en fonction de la manière dont le processus originel l'activera), et nous pouvons obtenir automatiquement la réponse la plus adaptée à la situation (la plus adaptée dans le sens d'une cohérence interne, si la réponse la plus adaptée pour le système est de faire une erreur, celui-ci pourra faire une erreur d'attribution cohérente avec sa manière de traiter l'information : ce qui est caractéristique de l'affordance). En effet, un des problèmes majeurs d'une conception représentationnaliste de la mémoire (qu'elle soit cognitiviste ou connexionniste) est que la structure de traitement est prédéterminée, elle est généralement faite de manière à ce que le modèle soit capable d'effectuer une tâche précise. Tandis que pour une mémoire

énonciviste, la manière de traiter l'information est apprise en fonction de l'histoire du système, ce n'est pas un système spécialisé mais qui se spécialise lui-même en fonction des situations rencontrées. De même, l'adaptation constante à l'environnement lui permet d'utiliser ses connaissances passées de manière à créer une nouvelle manière de créer de l'information. Cette caractéristique est très bien imagée par le raisonnement syllogistique d'une émergence en cascade : le système énonciviste est capable de créer de nouvelles covariances entre différentes covariances et de s'en servir comme nouvelle manière de traiter l'information. C'est une capacité nécessaire pour pouvoir reproduire le fonctionnement de la mémoire humaine.

La deuxième série de simulations portant sur la réplication des capacités de MINERVA2 se décline en quatre simulations : la capacité à compléter des informations manquantes selon le nombre de répétitions (rappel indicé), l'aspect épisodique d'une mémoire sans représentation (FURIG), la réaction à la nouveauté, et enfin les erreurs de recombinaisons illusives (MCE).

Le rappel indicé est une fonction de base de la mémoire humaine et ne dépend pas du paradigme expérimental dans lequel il est étudié. Nous avons pu montrer que ATHENA était capable de reproduire le comportement de MINERVA2 (capacité de rappel et apprentissage par répétition). Cependant, alors que la répétition augmente le contenu représentationnel à rappeler pour MINERVA2, pour ATHENA cette répétition renforce le processus appris de manière à ce qu'il soit plus efficace. Cette amélioration dans son efficacité dépend de ce que ATHENA est déjà capable de faire, et la courbe d'apprentissage suit une augmentation logarithmique, plus proche du comportement des humains que l'augmentation linéaire du modèle MINERVA2. En ce sens ATHENA parvient à reproduire un apprentissage plus proche de la mémoire humaine que son équivalent représentationniste. La simulation de l'expérience FURIG était nécessaire pour s'assurer que ATHENA ne perdait pas ses capacités épisodiques en apprenant non pas une représentation d'un épisode, mais en modifiant les processus déjà appris et distribués sur l'ensemble de ses traces. Intuitivement, nous pourrions considérer que de cette manière, le prototype des informations apprises devrait prendre le dessus sur les exemplaires. Mais il n'en est rien puisque ATHENA reproduit le pattern de fonctionnement de MINERVA2. La simulation de la réaction à la nouveauté nous a permis à la fois de nous assurer que pour ATHENA la nouveauté était plus surprenante que ce que la prototypie est familière (concept nécessaire dans une vision évolutionniste de la mémoire), mais aussi que l'emploi de valeurs flottantes était efficacement gérée par ATHENA. Le fait de ne devoir employer que des valeurs entières est en effet une limite du modèle MINERVA2 car dans les faits une composante n'est que rarement soit complètement activée, soit complètement inhibée, soit complètement inactive. Il existe toute une palette d'activations relatives qui est par ailleurs nécessaire pour l'apprentissage de covariances (le fait qu'une composante soit plus activée qu'une

autre est important). Enfin, la simulation des erreurs de recombinaisons illusoires était un complément nécessaire à la simulation du paradigme DRM. En effet, les effets DRM et MCE sont spécifiques à la construction de faux souvenirs mais sont issus de deux particularités élémentaires et complémentaires de la mémoire : l'interaction inter-traces et l'interaction inter-composantes. Bien que ces effets pourraient être considérés comme provenant d'une recombinaison de contenus mnésiques, le fait que cette simulation donne sur ATHENA des résultats similaires aux données expérimentales montre que ces effets ne sont pas incompatibles avec une mémoire de processus.

Cependant, ces résultats sont à modérer par la perte d'efficacité d'ATHENA comparativement à MINERVA2. En effet, une conséquence directe de l'emploi de covariances est que notre modèle ne se base que sur des relations de corrélation. Les corrélations sont efficaces pour approfondir les relations possibles entre les situations rencontrées, mais elles ne cherchent pas à reproduire des données objectivables (contrairement aux modèles représentationnalistes). ATHENA se trompe plus souvent que MINERVA2. Les mots anciens sont moins souvent considérés comme anciens, et les mots nouveaux sont moins souvent considérés comme nouveaux. Cette limite est due à l'absence d'un processus d'auto-évaluation de la pertinence d'une réponse. La fluence d'une situation n'est pas évaluée en fonction des fluences déjà obtenues dans des situations similaires. Dans la mesure où le modèle ATHENA n'utilise sa fluence de traitement (l'intensité) que pour affiner la création de nouvelles instances (d'ATHENA, au cœur des traces), la fluence n'est pas apprise, elle ne peut pas faire l'objet d'une fluence relative aux expériences passées. Par exemple, si la fluence du traitement d'un mot ancien est faible, cette non-fluence est non-fluente relativement à l'apprentissage du mot ancien. Une non-fluence non-fluente est pourtant censé améliorer la fluence du mot ancien (ce qui provoquerait moins d'erreurs d'attribution). De même, si un mot nouveau est « surprenamment » fluent, cette surprise devrait baisser la fluence du mot nouveau et limiter sa fausse reconnaissance. L'intégration de la fluence comme composante cognitive devrait pouvoir résoudre ce problème (nous y reviendrons).

La troisième série de simulation portant sur le comportement temporel se décline en quatre simulations : la cohérence de l'amorçage (interférence), la courbe de primauté-récurrence (serial position effect), l'interférence proactive (réverbération mnésique), et l'interférence rétroactive (prédiction mnésique).

La simulation d'interférence a permis de s'assurer que l'amorçage temporel présenté précédemment n'était pas uniquement dû à une répétition de l'item connu ou à une baisse de reconnaissance de la cible incongruente, mais bien à une interférence cohérente entre l'amorce et la cible. Une amorce congruente améliore la reconnaissance tandis qu'une amorce incongruente diminue la reconnaissance. Le lien existant entre deux situations qui se suivent n'est pas pris en compte par MINERVA2 (qui n'apprend pas la manière

dont il traite l'information), mais se retrouve dans ATHENA, introduisant la cohérence de ses caractéristiques temporelles. La simulation de l'effet de primauté-récence a permis de mettre en avant le fait que la courbe qui la caractérise n'est pas nécessairement due à une interférence entre modules de mémoires, ni même entre les représentations apprises des modèles à système unique. Cette courbe se retrouve également dans une mémoire énacliviste sans représentation où ce n'est pas le contenu représentationnel qui se modifie mais la manière de construire le souvenir. La simulation de l'effet de réverbération mnésique nous a permis d'expliquer comment un tel type de mémoire était capable de reproduire l'effet de primauté. Les premiers souvenirs définissent comment les suivants se construisent, et ce avec une intensité relative à la distance qui les sépare. Par exemple, dans une liste de mots, l'apprentissage du second mot sera construit grâce à l'interprétation qu'il est possible d'en faire avec le souvenir du premier mot ; tandis que le troisième mot sera construit à la fois grâce au premier et au second mot, ce qui limitera l'influence de la réverbération du premier mot dans le troisième comparativement au second. La construction temporelle des souvenirs n'est pas due à une temporalité absolue sur laquelle elle se baserait, mais sur la simple logique constructiviste de la création mnésique : un « temporalisme » subjectif. La simulation de l'effet de prédiction mnésique suit la même logique mais dans l'autre sens. En effet, la construction d'un nouveau souvenir demande aux traces une interprétation de la situation qui va les modifier en retour, de telle manière que les dernières traces sont moins modifiées que les précédentes (provoquant une interférence rétroactive dont l'intensité est relative à la distance temporelle qui sépare les souvenirs). La temporalité d'ATHENA est une construction directement issue de sa capacité à modifier les traitements appris. Tandis que MINERVA2 est, sans équivoque, incapable de réaliser un comportement temporel sans l'ajout d'un module dédié.

L'aspect fractal de la contextualisation des covariances les rendent dynamiques et permettent la construction d'une logique temporelle cohérente qui serait théoriquement suffisante pour qu'un système cherchant à se définir en tant que « soi » puisse percevoir des covariances temporelles sur lesquelles se baser. À cette fin, nous pouvons remarquer que l'interférence proactive est plus marquée que l'interférence rétroactive, ce qui est nécessaire pour la construction d'une flèche du temps cohérente avec la notre (i.e., du passé vers le futur). Mais cette particularité est aussi une limite du modèle ATHENA si l'on cherche à lui faire apprendre l'origine d'une action. En effet, l'origine d'une action sera moins fortement corrélée à celle-ci que sa conséquence, et ce type de simulation serait fortement biaisé. De même, en tentant de faire apprendre au modèle une suite de conséquences d'action, celui-ci ne pourra pas faire de distinction temporelle entre les couples d'items, et la conséquence d'une action modifiera l'action suivante de telle manière que les interactions entre action et conséquence de l'action seront biaisées. Nous avons rencontré ce problème en tentant de modéliser l'expérience de Brouillet, Vagnot, Milhau, Brunel, Briglia, Versace & Rousset (2015c) sur la conséquence de l'action, et ce problème ne serait résoluble qu'en

implémentant un mécanisme de distinction temporelle basé sur les ruptures de fluence (donc en permettant au modèle d'apprendre les évolutions de fluence).

Nous avons par ailleurs pu voir dans ces simulations que les modèles de type MINERVA2 comportaient deux insuffisances majeures. En premier lieu, il est impossible d'effectuer un rappel libre car il est nécessaire de présenter une sonde pour avoir un écho. Bien entendu, un rappel libre n'est jamais libre de consigne, mais il serait nécessaire de pouvoir produire plusieurs échos différents suite à la présentation d'une seule sonde. Deuxièmement, les modèles ATHENA et MINERVA2 sont tellement sensibles à l'ajout de nouvelles situations qu'il nous a fallu leur faire passer les tâches de reconnaissance en mode test (sans qu'ils apprennent la reconnaissance qu'ils viennent de faire). Ce n'est pas un problème pour les simulations que nous avons présentées, mais cette particularité limite les simulations possibles. Cette haute sensibilité s'explique simplement par le fait que ces modèles soient « jeunes » (i.e., la simulation est l'ensemble de l'historique des modèles), contrairement à un humain qui a dû se construire sur un intervalle de temps conséquent avant de pouvoir faire une tâche de reconnaissance. Ce problème est d'autant plus gênant que le modèle ATHENA construit sa manière d'interpréter les choses, il faudrait alors lui laisser plus de temps pour qu'il puisse se stabiliser. Malheureusement, la complexité mathématique d'un modèle fractal est telle que la puissance de calcul nécessaire à le faire tourner augmente de manière exponentielle au cours des situations rencontrées. Il n'existe virtuellement aucun ordinateur capable de le faire tourner à long terme. Ces problèmes nous poussent à envisager une modélisation en réseaux de neurones d'une mémoire énaïvistique de type ATHENA avant d'entamer une quelconque modification de son fonctionnement. Un réseau de neurones nous permettra d'utiliser une évolution fractale émulée par une distribution des covariances, nécessitant ainsi une puissance informatique conséquente mais stable dans le temps.

Les simulations présentées dans ce travail ont pu montrer qu'une mémoire multi-traces sans représentation conservait les mêmes patterns comportementaux qu'une mémoire multi-traces représentationnaliste, en plus de quoi l'absence de représentation permet une meilleure gestion de l'abstraction et un comportement temporel émergent. Bien que l'utilisation de représentations peut sembler intuitivement plus à même de modéliser le comportement de la mémoire, dans les faits les apports de l'énaïvisme radical permettent une meilleure adaptabilité de celle-ci à son environnement. La modélisation d'un tel type de mémoire est bien plus complexe que ce qui se fait habituellement, tant en termes théoriques qu'informatiques. Pourtant, une mémoire énaïvistique repose sur un nombre restreint de postulats, et ne rencontre toujours pas — à notre connaissance — d'incohérence interne.

Un modèle de mémoire est par définition une représentation du fonctionnement de la mémoire : c'est une simplification utile pour expliquer les théories qui lui ont donné

naissance, mais c'est surtout une construction issue de l'interprétation humaine, un ersatz. Pour pouvoir interagir avec ATHENA nous sommes contraints d'utiliser des sondes et des échos qui représentent des informations (mots, couleurs, etc.), mais pour ATHENA ces vecteurs sont directement son environnement sensorimoteur. Pour le modèle, ces vecteurs n'ont aucune signification, et comme il n'apprend pas ces informations telles quelles (mais la manière dont il parvient à les interpréter), il leur donne du sens en fonction de son vécu. Le fait que nous employons des vecteurs représentationnels pour pouvoir modifier et analyser son comportement ne change en rien le fait que celui-ci n'apprenne pas de représentations. Malgré cette contrainte technique, ATHENA reste fondamentalement un modèle énaïviste, et par conséquent, un modèle de mémoire sans représentation.

Le simple fait que notre modèle soit capable de manipuler des abstractions émergentes sans représentation démontre sa capacité à effectuer de manière cohérente des traitements de haut niveau uniquement à partir de processus de bas niveau. Cette dissociation historique entre processus de haut et bas niveau n'est pas une nécessité, et employer l'énaïvisme comme base explicative des processus de bas niveau est suffisante pour se passer de représentations dans l'explication des processus de haut-niveau. L'énaïvisme conservateur qui tente de concilier les apports de l'énaïvisme avec la culture cognitiviste est — au mieux — superflu. Conformément aux travaux de Varela (1993, 1996), l'énaïvisme n'a de pertinence que dans un paradigme qui rejette en bloc la notion de représentation. L'énaïvisme est inévitablement un processus propre à l'énaïvisme radical.

4.4 Futurs travaux

L'énaïvisme est souvent critiquée pour son formalisme limité et son manque d'applicabilité. Dans ce travail, le modèle ATHENA tente modestement d'apporter une pierre à l'édifice énaïviste en rendant possible sa mise en situation par des protocoles expérimentaux. Cependant, bien que les apports du modèle ATHENA soient intéressants, ils restent encore insuffisants. Il est un apport du modèle Act-In qui — de par sa complexité — n'a pas été pris en compte dans ce travail, mais qui pourrait théoriquement régler les problèmes rencontrés par notre modèle sur les erreurs d'attribution : l'intégration des émotions. En effet, une des particularités du modèle Act-In est de concevoir les émotions comme une composante cognitive au même titre que les composantes sensorimotrices, mais capable d'améliorer l'intégration (Versace et al., 2014). Il s'agit d'une particularité nécessaire dans une mémoire énaïviste, laquelle se base exclusivement sur l'expérience vécue pour construire une manière de traiter les informations. En ce sens, et puisqu'il n'est aucunement nécessaire de faire de distinction entre traitement de haut et bas niveau (seul le niveau de complexité mathématique du traitement change), nous entendons par émotions le traitement émotionnel à son niveau le plus fondamental : les covariances sensorimotrices. En effet, n'oublions pas que le traitement cognitif fait partie de l'envi-

ronnement pour un système énonactif, donc les covariances sensorimotrices prennent également en compte la perception et la modification des traitements internes. Dès lors, l'émotion dans un paradigme énonactif est fondamentalement un processus d'évaluation des traitements effectués et de leur fluence (voir Brouillet et al., 2015b ; Milhau et al., 2015). ATHENA comporte déjà un tel processus d'évaluation : l'intensité, qui — de par son lien avec la notion de familiarité — est une évaluation de la fluence du traitement, et concorde avec la notion de fluence comme indice méta-cognitif (Oppenheimer, 2008 ; Alter & Oppenheimer, 2009). Cependant, comme nous avons pu le voir, cette évaluation de fluence n'est pas apprise et n'a d'influence que sur la construction du souvenir, pas sur la production de l'écho auquel il est associé. Implémenter le traitement émotionnel comme une composante cognitive devra se faire sous la forme d'une fluence apprise.

La fluence est fondamentalement une information d'évaluation du système cognitif, dès lors, en apprenant et en utilisant la fluence comme une composante cognitive, nous retrouvons au cœur de la cognition un traitement propre à l'évaluation de cette même cognition. Si la fluence est une composante cognitive elle est alors un processus cognitif qui évalue le traitement cognitif : une méta-cognition. Cette méta-cognition étant une composante cognitive au même titre qu'une composante sensorimotrice, la fluence est alors également une composante méta-cognitive. De même, si la composante méta-cognitive est une composante cognitive elle sera également une composante méta-méta-cognitive, et ainsi de suite. Intégrer la fluence au cœur d'une mémoire à système unique permet ainsi de résoudre les évaluations de méta-cognition en cascade sans avoir besoin de faire appel à une modularité méta-cognitive spécifique (Servajean, 2016 ; Servajean, Briglia & Brouillet, 2017). En intégrant une composante de fluence apprise dans une modélisation de l'écphorie synergétique, l'appariement global des fluences de traitement à la situation doit permettre une modification de l'état de fluence du traitement de la situation. Une thèse portant l'aspect système unique de la méta-cognition est en cours dans l'équipe de recherche (celle de P. Servajean). De cette manière, la production sensorimotrice ne dépend plus uniquement de la contextualisation des covariances mais également de l'évaluation de cette contextualisation en fonction des évaluations passées. Cette caractéristique est essentielle pour résoudre les erreurs d'attribution produites par le modèle ATHENA. Par exemple, si la reconnaissance d'un mot ancien produit une non-fluence, plutôt que de se limiter à juger le mot comme peu familier et provoquer une omission, l'évaluation de cette non-fluence peut-être évaluée comme non-fluente (parce que le mot rappelle l'environnement familier de la tâche de reconnaissance). Cette non-fluence non-fluente gêne la non-fluidité de l'évaluation du traitement et la fluence du traitement global s'en trouvera améliorée (et le mot sera mieux reconnu). De même, si la reconnaissance d'un mot nouveau provoque une fluence qui est évaluée comme non-fluente, sa fluence s'en trouvera amoindrie et la fausse reconnaissance sera atténuée. L'émotion est une modification dans l'interprétation (pour un humain, la fluence est agréable et la non-fluence désagréable : ce qui est agréable n'a pas

besoin d'être modifié, ce qui est désagréable pousse à la modification d'une attribution). Cette conception théorique où la fluence facilite le traitement et la non-fluence l'entrave est parfaitement compatible avec le paradigme énaïviste et ses évaluations de covariances. Il s'agit même d'une des bases de l'autopoïèse où un système énaïviste ne cherche qu'à se stabiliser. Le rôle de la fluence de traitement est essentiel dans la stabilisation d'un système complexe : la fluence est caractéristique d'un fonctionnement efficace et ne va que renforcer le comportement qui lui a donné naissance, tandis que la non-fluence est caractéristique d'un fonctionnement inadapté qui va modifier le comportement qui lui a donné naissance en sens inverse. C'est en ce sens que la fluence est un bon indicateur de l'état émotionnel d'un système autopoïétique, et c'est grâce à la fluence comme composante cognitive que nous parviendrons à améliorer le comportement du modèle ATHENA.

En définitive, la méta-cognition a comme utilité de faciliter la modification d'un système autopoïétique afin de tendre vers un état de stabilité de manière plus efficace. L'interprétation des modifications méta-cognitives permet alors d'associer un état de fluence perçue à un état de fluence optimale. Cette association permettrait théoriquement à un système autopoïétique d'être influencé par un état de stabilité hypothétique jugé plus fluent que l'état actuel du système : l'anticipation d'un état de stabilité vers lequel tendrait le système. L'anticipation est ici différente du comportement de prédiction temporelle que nous avons pu voir dans la partie simulation. En effet, la prédiction temporelle ne fait que prévoir la situation qu'il serait probable de percevoir, tandis que l'anticipation prévoit la situation qu'il serait pertinent de percevoir pour le système. Les apports théoriques d'un tel système d'anticipation de fluence stabilisée devraient conséquemment faire valoir la pertinence du paradigme énaïviste. En effet, l'énaïvisme stipule que la construction de la signification s'effectue en fonction de l'expérience vécue, donc de l'expérience consciente. Nous avons prêté une attention particulière dans ce travail à ne pas créer d'incohérence interne entre la construction de la mémoire et la faculté d'un système cognitif à adopter une conscience positionnelle (conformément à la phénoménologie, la conscience est toujours conscience de quelque chose).

Sortir de la notion de « soi » comme un absolu prédéfini nous a permis de concevoir un modèle se passant efficacement de représentations, mais ce principe permet aussi d'envisager un système autopoïétique capable d'intentionnalité. Les émotions ne sont pas en soi une anticipation intentionnelle, mais elles la permettraient si elles étaient considérées comme un indice méta-cognitif. Puisque la fluence est à la fois une composante cognitive et méta-cognitive, le jeu infini du double miroir Sartrien (en-soi/pour-soi) pourrait en émerger afin de permettre la construction d'un positionnement intentionnel.

4.5 Conclusion

À travers ce travail, nous avons tenté de donner à l'énaclivisme radical quatre paramètres fondamentaux sur lesquels modéliser une mémoire sans représentation. La formalisation mathématique d'une mémoire énacliviste n'avait — à notre connaissance — jamais été réalisée, encore moins testée expérimentalement. En conclusion, un tel système permet une meilleure gestion de l'abstraction et un comportement temporel émergent, mais un cinquième paramètre fondamental traitant de la fluence comme indice méta-cognitif manque à l'appel.

Les difficultés matérielles d'un temps de calcul exponentiel, et les limitations inhérentes à un système simplifié de type MINERVA2 nous poussent à envisager une modélisation des théories développées dans ce travail sous la forme d'un réseau de neurones (et d'un environnement adéquat pour sa construction auto-organisée). Le traitement émotionnel manquant à ATHENA devrait être modélisable par l'emploi de la fluence comme composante cognitive (et donc comme indice méta-cognitif). Si la méta-contextualisation de covariances sensorimotrices peut être modélisée dans un réseau de neurones, la fluence méta-cognitive peut — quant à elle — être modélisée par un réseau d'atocytes (interagissant par vagues calciques). L'interdépendance de ces deux réseaux au niveau biologique fait déjà l'objet de nombreuses recherches (pour une revue, cf., Perea et al., 2014) et est directement liée à l'efficacité mnésique, au comportements volontaires ainsi qu'aux traitements émotionnels (pour une revue, cf., Pereira & Furlan, 2010). Un tel réseau devrait permettre à une mémoire énacliviste d'ajuster ses erreurs d'attribution et de faire émerger une anticipation intentionnelle.

Bibliographie

- Agre, P. E. and Chapman, D. (1987). Pengi : An implementation of a theory of activity. In *AAAI*, volume 87, pages 286–272.
- Alter, A. L. and Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and social psychology review*, 13(3) :219–235.
- Anderson, J. R. and Pirolli, P. L. (1984). Spread of activation. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 10(4) :791.
- Atkinson, R. C. and Shiffrin, R. M. (1968). Human memory : A proposed system and its control processes. *Psychology of learning and motivation*, 2 :89–195.
- Bach-y Rita, P. (1967). Sensory plasticity. *Acta Neurologica Scandinavica*, 43(4) :417–426.
- Barbaras, R. (1994). La perception—essai sur le sensible, paris : Hatier. *Coll. Optiques*.
- Benjamin, A. S., Bjork, R. A., and Hirshman, E. (1998). Predicting the future and reconstructing the past : A bayesian characterization of the utility of subjective fluency. *Acta Psychologica*, 98(2) :267–290.
- Bills, A. G. (1934). *General experimental psychology*. Longmans, Green and Co.
- Bitzer, S., Park, H., Blankenburg, F., and Kiebel, S. J. (2014). Perceptual decision making : drift-diffusion model is equivalent to a bayesian model. *Frontiers in human neuroscience*, 8.
- Briglia, J., Servajean, P., Michalland, A.-H., Brunel, L., and Brouillet, D. (2017). Modeling an enactivist multiple-trace memory. athena : a fractal model of human memory. *Manuscrit en révision*.
- Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In Rosch, E. and Lloyd, B., editors, *Cognition and categorization*, pages 3–170.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47 :139–159.
- Brouillet, D., Brouillet, T., Milhau, A., Heurley, L., Vagnot, C., and Brunel, L. (2015a). Word-to-picture recognition is a function of motor components mappings at the stage of retrieval. *International Journal of Psychology*, 51(5) :397-402.

- Brouillet, D. and Jimenez, M. (2007). Document interne au laboratoire.
- Brouillet, D., Milhau, A., and Brouillet, T. (2015b). When “good” is not always right : effect of the consequences of motor action on valence-space associations. *Frontiers in psychology*, 6.
- Brouillet, D., Vagnot, C., Milhau, A., Brunel, L., Briglia, J., Versace, R., and Rousset, S. (2015c). Sensory–motor properties of past actions bias memory in a recognition task. *Psychological research*, 79(4) :678–686.
- Bruner, J. S. (1966). *Toward a theory of instruction*, volume 59. Harvard University Press.
- Busemeyer, J. R. and Wang, Z. (2015). What is quantum cognition, and how is it applied to psychology? *Current Directions in Psychological Science*, 24(3) :163–169.
- Carbonnel, S., Charnallet, A., and Moreaud, O. (2010). Organisation des connaissances sémantiques : des modèles classiques aux modèles non abstraits. *Revue de neuropsychologie*, 2(1) :22–30.
- Chemero, A. (2013). Radical embodied cognitive science. *Review of General Psychology*, 17(2) :145.
- Chomsky, N. (1969). *La linguistique cartésienne*. Ed. du seuil.
- Chu, S. (2007). Olfactory conditioning of positive performance in humans. *Chemical senses*, 33(1) :65–71.
- Clark, A. (1997). The dynamical challenge. *Cognitive Science*, 21(4) :461–481.
- Clark, A. and Grush, R. (1999). Towards a cognitive robotics. *Adaptive Behavior*, 7(1) :5–16.
- Clark, A. and Toribio, J. (1994). Doing without representing? *Synthese*, 101(3) :401–431.
- Cosmelli, D. and Thompson, E. (2010). Embodiment or envatment? reflections on the bodily basis of consciousness. *Enaction : Towards a new paradigm for cognitive science*, pages 361–385.
- Cretenet, J. and Dru, V. (2009). Influence of peripheral and motivational cues on rigid–flexible functioning : Perceptual, behavioral, and cognitive aspects. *Journal of Experimental Psychology : General*, 138(2) :201.
- Damas, L., Mille, A., and Versace, R. (2002). Prendre en compte les comportements cognitifs des apprenants dans la conception de systèmes d’assistance à l’apprentissage humain. *TICE 2002*, pages 111–118.

- Damasio, A. R. (1994). *Erreur de Descartes (L')*. Odile Jacob.
- Davidoff, J., Davies, I., and Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature*, 398(6724) :203.
- Dennett, D. (1978). Beliefs about beliefs. *Behav. Brain Sci.*, 1 :568–570.
- Dennett, D. (1984). *Elbow room : The varieties of free will worth wanting*. MIT Press.
- Descartes, R. (1641). *Méditations métaphysiques*. Nathan, Paris, 2004.
- Dewar, M. T., Cowan, N., and Della Sala, S. (2007). Forgetting due to retroactive interference : A fusion of müller and pilzecker’s (1900) early insights into everyday forgetting and recent research on anterograde amnesia. *Cortex*, 43(5) :616–634.
- Di Paolo, E. A., Rohde, M., and De Jaegher, H. (2010). Horizons for the enactive mind : Values, social interaction, and play. *Enaction : Toward a new paradigm for cognitive science*, pages 33–87.
- Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events : a neurophysiological study. *Experimental brain research*, 91(1) :176–180.
- Ebbinghaus, H. (1913). *Memory : A contribution to experimental psychology*. Number 3. University Microfilms.
- Einstein, A. and Infeld, L. (1948). *L'évolution des idées en physique*. Flammarion Paris.
- Ferrier, L., Streff, A., Martin, S., Brouillet, D., Barkat-Defradas, M., and Jiménez, M. (2009). Influence de stimuli olfactifs dans une tâche d'évaluation hédonique de couleurs : les yeux voient ce que le nez sent. *L'Année psychologique*, 109(03) :361–381.
- Fisher, G. H. (1967). Preparation of ambiguous stimulus materials. *Attention, Perception, & Psychophysics*, 2(9) :421–422.
- Fodor, J. A. (1981). *Representations : Philosophical essays on the foundations of cognitive science*. MIT Press.
- Fodor, J. A. (1987). *Psychosemantics : The problem of meaning in the philosophy of mind*, volume 2. MIT press.
- Franklin, D. R. and Mewhort, D. (2015). Memory as a hologram : an analysis of learning and recall. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 69(1) :115.

- Fuster, J. M., Alexander, G. E., et al. (1971). Neuron activity related to short-term memory. *Science*, 173(3997) :652–654.
- Gibson, E. J. (1982). The concept of affordances in development : The renaissance of functionalism. In *The concept of development : The Minnesota symposia on child psychology*, volume 15, pages 55–81. Lawrence Erlbaum Hillsdale, NJ.
- Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron*, 14(3) :477–485.
- Goldstein, K. (1934). *La structure de l'organisme : introduction à la biologie à partir de la pathologie humaine (E. Burckhardt & J. Kuntz, trads.)*. Paris : Gallimard.
- Greene, B. (2005). *L'espace, le temps, la réalité : tout est à repenser*. Gallimard.
- Grush, R. (2004). The emulation theory of representation : Motor control, imagery, and perception. *Behavioral and brain sciences*, 27(3) :377–396.
- Haken, H. (1983). *Synergetics : An introduction*. Springer-Verlag Berlin Heidelberg.
- Hansen, T., Olkkonen, M., Walter, S., and Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature neuroscience*, 9(11) :1367.
- Havelange, V., Lenay, C., and Stewart, J. (2003). Les représentations : mémoire externe et objets techniques. *Intellectica*, 35 :115–131.
- Hawking, S. (1993). *Une brève histoire du temps*. J'ai lu.
- Hebb, D. O. (1949). *The organization of behavior : A neuropsychological approach*. John Wiley & Sons.
- Heidegger, M. (1927). *Sein und zeit*. Halle ad S. : M. Niemeyer.
- Heurley, L. P., Brouillet, T., Chesnoy, G., and Brouillet, D. (2013). Color perception involves color representations firstly at a semantic level and then at a lexical level. *Cognitive processing*, 14(1) :19–29.
- Hintzman, D. L. (1984). Minerva 2 : A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, 16(2) :96–101.
- Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review*, 93(4) :411–428.
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95(4) :528.
- Hutto, D. and Myin, E. (2012). *Radicalizing enactivism : Basic minds without content*. Mit Press.

- Hutto, D. D. (2005). Knowing what? radical versus conservative enactivism. *Phenomenology and the Cognitive Sciences*, 4(4) :389–405.
- Jackendoff, R. (1987). *Consciousness and the computational mind*. The MIT Press.
- James, W. (1890). *The principles of psychology*. New York, NY : Henry Holt.
- Klein, É. (2009). *Les tactiques de Chronos*. Flammarion.
- Leboe-McGowan, J. P. and Whittlesea, B. W. (2013). 17 through the scape looking glass—sources of performance and sources of attribution. *The Oxford Handbook of Cognitive Psychology*, page 243.
- Maier, V. and Moore, R. K. (2005). An investigation into a simulation of episodic memory for automatic speech recognition. In *Ninth European Conference on Speech Communication and Technology*.
- Martins, M. D., Gingras, B., Puig-Waldmueller, E., and Fitch, W. T. (2017). Cognitive representation of “musical fractals” : Processing hierarchy and recursion in the auditory domain. *Cognition*, 161 :31–45.
- Mattei, T. A. (2014). Unveiling complexity : non-linear and fractal analysis in neuroscience and cognitive psychology. *Frontiers in computational neuroscience*, 8.
- Maturana, H. R. and Varela, F. J. (1987). *The tree of knowledge : The biological roots of human understanding*. New Science Library/Shambhala Publications.
- Maturana, H. R., Varela, F. J., and Jullien, F.-C. (1994). *L'arbre de la connaissance*. Ed. Addison-Wesley France.
- McCrary Jr, J. W. and Hunter, W. S. (1953). Serial position curves in verbal learning. *Science*, 117(3032) :131–4.
- Medin, D. L. and Ortony, A. (1989). Psychological essentialism. *Similarity and analogical reasoning*, 179.
- Menary, R. (2010). Introduction to the special issue on 4e cognition. *Phenomenology and the Cognitive Sciences*, 9(4) :459–463.
- Merleau-Ponty, M. (1942). *La structure du comportement*. Presses universitaires de France.
- Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Éditions Gallimard.
- Milhau, A., Brouillet, T., and Brouillet, D. (2015). Valence–space compatibility effects depend on situated motor fluency in both right-and left-handers. *The Quarterly Journal of Experimental Psychology*, 68(5) :887–899.

- Miller, E. K., Erickson, C. A., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of neuroscience*, 16(16) :5154–5167.
- Millikan, R. G. (1995). Pushmi-pullyu representations. *Philosophical perspectives*, 9 :185–200.
- Miyashita, Y. and Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex yasushi miyashita. *Nature*, 331(6151) :68–70.
- Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, 319(5869) :1543–1546.
- Morin, R., DeRosa, D., and Stultz, V. (1967). Recognition memory and reaction time. *Acta Psychologica*, 27 :298–305.
- Murdock Jr, B. B. (1968). Serial order effects in short-term memory. *Journal of Experimental Psychology*, 76(4p2) :1.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory : Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of experimental psychology : general*, 106(3) :226.
- Onifer, W. and Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension : Effects of frequency of meaning and contextual bias. *Memory & Cognition*, 9(3) :225–236.
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in cognitive sciences*, 12(6) :237–241.
- Osth, A. F. and Dennis, S. (2015). Sources of interference in item and associative recognition memory. *Psychological review*, 122(2) :260.
- O’Connor, A. R. and Moulin, C. J. (2008). The persistence of erroneous familiarity in an epileptic male : Challenging perceptual theories of déjà vu activation. *Brain and Cognition*, 68(2) :144–147.
- O’Connor, A. R. and Moulin, C. J. (2010). Recognition without identification, erroneous familiarity, and déjà vu. *Current Psychiatry Reports*, 12(3) :165–173.
- Paivio, A. and Csapo, K. (1969). Concrete image and verbal memory codes. *Journal of Experimental psychology*, 80(2p1) :279.
- Perea, G., Sur, M., and Araque, A. (2014). Neuron-glia networks : integral gear of brain function. *Frontiers in cellular neuroscience*, 8.

- Pereira, A. and Furlan, F. A. (2010). Astrocytes and human cognition : modeling information integration and modulation of neuronal activity. *Progress in neurobiology*, 92(3) :405–420.
- Piaget, J. (1936). *La naissance de l'intelligence chez l'enfant*. Delachaux.
- Piaget, J. (1937). *La construction du réel chez l'enfant*. Delachaux & Niestlé.
- Piaget, J. (1970). *Psychologie et épistémologie*. Paris, Gonthier.
- Poincaré, H. (1902). *La science et l'hypothèse*. Flammarion, Paris.
- Port, R. F., Cummins, F., and McAuley, J. D. (1995). *Naive time, temporal patterns, and human audition*. The MIT Press.
- Prigogine, I. and Stengers, I. (1979). *La nouvelle alliance : métamorphose de la science*. Gallimard.
- Raaijmakers, J. G. and Shiffrin, R. M. (1981). Search of associative memory. *Psychological review*, 88(2) :93.
- Rasiwasia, N. (2017). 'red' strawberries dissected. <https://www.linkedin.com/pulse/case-red-strawberries-nikhil-rasiwasia>. Accessed : 01-10-2017.
- Reinitz, M. T., Lammers, W. J., and Cochran, B. P. (1992). Memory-conjunction errors : Miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition*, 20(1) :1–11.
- Rousset, S. (2000). Les conceptions "système unique" de la mémoire : aspects théoriques. *Revue de neuropsychologie*, 10(1) :27–51.
- Rudrauf, D., Lutz, A., Cosmelli, D., Lachaux, J.-P., and Le Van Quyen, M. (2003). From autopoiesis to neurophenomenology : Francisco varela's exploration of the biophysics of being. *Biological research*, 36(1) :27–65.
- Sartre, J. (1936). *L'imagination*. Paris, PUF.
- Sartre, J.-P. (1943). *L'être et le néant, essai d'ontologie phénoménologique*. Gallimard Paris.
- Schwartz, B. L., Benjamin, A. S., and Bjork, R. A. (1997). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science*, 6(5) :132–137.
- Servajean, P. (2016). Approche système unique de la métacognition. 57ème congrès sfp 2016, 07-09 septembre, paris.

- Servajean, P., Briglia, J., and Brouillet, D. (2017). What a surprising surprise : a unique level model of (meta)cognition. *Manuscrit en cours de préparation*.
- Smith, L. B. and Thelen, E. E. (1993). A dynamic systems approach to development : Applications. In *This book grew out of a workshop, " Dynamic Systems in Development, " held for the Society for Research in Child Development in Kansas City, KS, Apr 1989*. The MIT Press.
- Smolensky, P. (1992). Ia connexionniste, ia symbolique et cerveau. *Introduction aux sciences cognitives*, pages 79–106.
- Smuts, J. C. (1926). *Holism and evolution*. New York : The Macmillan company.
- Swinney, D. A. (1979). Lexical access during sentence comprehension :(re) consideration of context effects. *Journal of verbal learning and verbal behavior*, 18(6) :645–659.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind : Statistics, structure, and abstraction. *science*, 331(6022) :1279–1285.
- Thompson, E. (2007). *Mind in Life : Biology, Phenomenology and the Sciences of Mind*. Cambridge, MA : Harvard University Press.
- Tiberghien, G. (1997). *La mémoire oubliée*. Editions Mardaga.
- Tulving, E. (1974). Cue-dependent forgetting : When we forget something we once knew, it does not necessarily mean that the memory trace has been lost ; it may only be inaccessible. *American Scientist*, 62(1) :74–82.
- Tulving, E. (1995). Organization of memory : Quo vadis. *The cognitive neurosciences*, pages 839–847.
- Tulving, E., Le Voi, M., Routh, D., and Loftus, E. (1983). Ecphoric processes in episodic memory [and discussion]. *Philosophical Transactions of the Royal Society of London B : Biological Sciences*, 302(1110) :361–371.
- Turvey, M. T., Shaw, R. E., Reed, E. S., and Mace, W. M. (1981). Ecological laws of perceiving and acting : In reply to fodor and pylyshyn (1981). *Cognition*, 9(3) :237–304.
- Varela, F. J. (1979). *Principles of biological autonomy*. Appleton & Lange.
- Varela, F. J. (1988). *Invitation aux sciences cognitives*. Seuil.
- Varela, F. J. (1995). The emergent self. *The Third Culture*, pages 209–22.
- Varela, F. J. (1996). Neurophenomenology : A methodological remedy for the hard problem. *Journal of consciousness studies*, 3(4) :330–349.

- Varela, F. J. (1997). A science of consciousness as if experience mattered. In *Towards the science of consciousness : The second Tucson discussions and debates*, pages 31–44. MIT Press.
- Varela, F. J. (1999). L’auto-organisation : vrai ou faux miracle?’. *Science et conscience*, page 107.
- Varela, F. J. (2002). Le présent spéculaire : une neurophénoménologie de la conscience du temps. *Naturaliser la phénoménologie, Paris, CNRS Éditions*.
- Varela, F. J., Maturana, H. R., and Uribe, R. (1974). Autopoiesis : the organization of living systems, its characterization and a model. *Biosystems*, 5(4) :187–196.
- Varela, F. J., Thompson, E., and Rosch, E. (1991). *The embodied mind : Cognitive science and human experience*. MIT press.
- Varela, F. J., Thompson, E., and Rosch, E. (1993). L’inscription corporelle de l’esprit. *Paris, Seuil*.
- Versace, R., Labeye, E., Badard, G., and Rose, M. (2009). The contents of long-term memory and the emergence of knowledge. *European journal of cognitive psychology*, 21(4) :522–560.
- Versace, R., Vallet, G. T., Riou, B., Lesourd, M., Labeye, É., and Brunel, L. (2014). Act-in : An integrated view of memory mechanisms. *Journal of Cognitive Psychology*, 26(3) :280–306.
- Werner, G. (2010). Fractals in the nervous system : conceptual implications for theoretical neuroscience. *Frontiers in physiology*, 1.
- Whittlesea, B. W. (1987). Preservation of specific experiences in the representation of general knowledge. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 13(1) :3.
- Whittlesea, B. W. (1993). Illusions of familiarity. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 19(6) :1235.
- Whittlesea, B. W. (1997). Production, evaluation, and preservation of experiences : Constructive processing in remembering and performance tasks. *The Psychology of Learning and Motivation*, 37 :211–264.
- Whittlesea, B. W. (2002). False memory and the discrepancy-attribution hypothesis : the prototype-familiarity illusion. *Journal of Experimental Psychology : General*, 131(1) :96.

- Whittlesea, B. W. and Leboe, J. P. (2000). The heuristic basis of remembering and classification : fluency, generation, and resemblance. *Journal of Experimental Psychology : General*, 129(1) :84.
- Whittlesea, B. W. and Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language*, 49(1) :62–79.
- Whittlesea, B. W. and Williams, L. D. (2000). The source of feelings of familiarity : the discrepancy-attribution hypothesis. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 26(3) :547.
- Whittlesea, B. W. and Williams, L. D. (2001). The discrepancy-attribution hypothesis : I. the heuristic basis of feelings and familiarity. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 27(1) :3.
- Wundt, W. (1874). *Principes de psychologie physiologique*. Paris, Alcan.

Code source du modèle ATHENA

A.1 Licence

Le logiciel ATHENA est un logiciel libre dont la licence — CeCILL-B — est issue de certaines propriétés du projet GNU GPL et est compatible avec les spécificités de la licence BSD, mais adaptée aux nécessités du droit français. Une copie du contrat de licence est disponible à l'adresse suivante :

http://www.cecill.info/licences/Licence_CeCILL-B_V1-fr.html

CeCILL a été créée par des organismes de recherche français : le CEA (Commissariat à l'énergie atomique), le CNRS (Centre national de la recherche scientifique) et l'INRIA (Institut national de recherche en informatique et en automatique) afin de garantir aux utilisateurs et aux créateurs de logiciels libres le respect du droit français en matière de responsabilité civile et de droit à la propriété intellectuelle.

Les spécificités de la licence CeCILL-B sont avant tout celles des logiciels libres : libre accès au code source, libre utilisation, libre modification, libre redistribution. L'acceptation du contrat de licence débute automatiquement et systématiquement à la date de téléchargement du logiciel par l'utilisateur. En contrepartie, la modification de ATHENA ou le rajout d'un module externe lié à une volonté de redistribuer la version modifiée s'accompagne d'une forte obligation de citation de l'auteur et du contrat de licence du logiciel d'origine, dans la nouvelle documentation (ou article), au sein de l'interface du logiciel ainsi qu'au cœur du média de redistribution ; et ce, de manière facilement accessible aux utilisateurs.

Cette licence permet une diffusion libre et gratuite des avancées scientifiques qui pourraient découler de l'utilisation de ce logiciel, tout en conservant les spécificités du droit à la propriété intellectuelle. Ce logiciel ne peut alors en aucun cas faire l'objet d'une valeur marchande, et l'auteur n'est légalement responsable en aucun cas des conséquences de l'utilisation de ce logiciel par des tiers.

ATHENA a été élaboré en Smalltalk sur la plateforme Pharo. Pharo est un système réflexif (un programme peut modifier sa structure et son comportement lors de l'exécution) avec un typage dynamique (il n'y a pas de structure de variables à l'utilisation — un type de variable se définit par son comportement pratique). Il est à la fois plate-forme

de programmation, environnement d'exécution de logiciel et le langage dans lequel il est programmé : il a, en effet, la particularité d'être écrit en lui même, ce qui le rend compatible avec n'importe quel ordinateur, et en fait l'outil informatique phénoménologique par excellence. Enfin, son principal avantage est de pouvoir être modifié pendant l'exécution d'un programme : la programmation, la compilation, l'exécution, et le debugger font partie du même environnement dynamique (on peut passer de l'un à l'autre à n'importe quel moment).

Outre ses avantages pratiques et son statut de logiciel libre, nous avons choisi Pharo pour sa facilité de diffusion dans la communauté scientifique. En effet, tout comme la culture, la recherche n'a de valeur que si elle est librement partagée. Ce qui est aussi un des enjeux de SmalltalkHub, plate-forme de gestion de projets (communautaires ou personnels) – et de stockage/partage de codes sources — appliquée à Smalltalk et Monticello. Le dépôt de ATHENA y est alors publiquement accessible, ce qui autorise (et encourage) toute utilisation, modification et distribution du logiciel.

<http://smalltalkhub.com/#!/~JohanBriglia/ATHENA/>

A.2 Code-source

```
Object subclass: #Athena
  instanceVariableNames: 'mnesicFields similarities intensity
    archetypicalities fluence sensorimotorResonance
    sensorimotorField emptyField newField gatheredField
    gatheringInterval canProcreate '
  classVariableNames: ''
  poolDictionaries: ''
  category: 'ATHENA'!
!Athena commentStamp: 'JohanBriglia 12/21/2016 17:06' prior: 0!
The Athena class is the main object of the model.
enumeration protocol refers to methods used to enumerate traces
    , sub-traces , etc.
gears refers to the main methods running the model core (
    Firstly see #inject:).
instance variables refers to different variables used in the
    model application (probe, echo, etc)
```

Many thanks to Camille Teruel and Nicholas Petton for their help
in refactoring the code (x196 faster)

Copr. Johan BRIGLIA 2016

johan.briglia@gmail.com

Athena is a computer program whose purpose is to modelize the dynamic nature of human memory according to radical enactivist theories in single-system multiple-traces memory modeling. Athena is an improvement of Hintzman's Minerva2 ideas (1984, 1986, 1988) and do not use any part of the original software.

Athena is governed by the CeCILL-B license under French law and abiding by the rules of distribution of free software. You can use, modify and/or redistribute the software under the terms of the CeCILL-B license as circulated by CEA, CNRS and INRIA at the following URL

<http://www.cecill.info>

As a counterpart to the access to the source code and rights to copy, modify and redistribute granted by the license, users are provided only with a limited warranty and the software's author, the holder of the economic rights, and the successive licensors have only limited liability.

In this respect, the user's attention is drawn to the risks associated with loading, using, modifying and/or developing or reproducing the software by the user in light of its specific status of free software, that may mean that it is complicated to manipulate, and that also therefore means that it is reserved for developers/researchers and experienced professionals having in-depth computer and psychological knowledge. Users are therefore encouraged to load and test the software's suitability as regards their requirements in conditions enabling the security of their systems and/or data to be ensured and, more generally, to use and operate it in the same conditions as regards security.

The fact that you are presently reading this means that you have

had knowledge of the CeCILL-B license and that you accept its terms.!

```
!Athena methodsFor: 'gears' stamp: 'JohanBriglia 12/16/2016
```

```
16:59'!
```

```
calculateFluence
```

```
"method used to calculate intensity"
```

```
similarities := mnesicFields collect: [ :each | each inject:
  sensorimotorField ].
```

```
"this recursive method calls inject: for each trace, and sets
  similarities values for each
see #AthenaMnesicField for further calculations"
```

```
intensity := similarities inject: 0 into: [ :sum :x | x abs +
  sum ].
```

```
"this is not the final intensity (called fluence), it sums all
  absolute similarities
in order to get P(X)"
```

```
archetypicalities := [similarities collect: [ :each | (each /
  intensity) asFloat ]] "activations as P(xi|X)"
```

```
on: ZeroDivide do: [ :x | similarities ].
```

```
"this happens only when all similarities equals zero"
```

```
fluence := (1 to: mnesicFields size) inject: 0 into: [ :f :x |
  (similarities at: x) abs * (archetypicalities at: x) + f ].
```

```
"final intensity : sums all absolute similarities multiplied
  by their respective activations"! !
```

```
!Athena methodsFor: 'gears' stamp: 'JohanBriglia 12/16/2016
```

```
16:45'!
```

```
calculateResonance
```

```
"calculate the echo
```

```
as the sum of all contextual traces content multiplied with
  their respective activation"
```

```

sensorimotorResonance :=
  (1 to: mnesicFields size)
  inject: emptyField copy
  into: [ :field :x |
    (mnesicFields at: x) resonance * (archetypicalities at: x) +
    field ]! !

!Athena methodsFor: 'gears' stamp: 'JohanBriglia 12/16/2016
  16:50'!
calculateNewField

"returns the trace as gathered values in collections"

newField := sensorimotorResonance + sensorimotorField / 2.
"the new trace values are the mean between the probe ones and
  the echo ones"

^canProcreate "if it can create new Athena instances (number of
  features > 2)"
ifTrue: [ self gatherNewField ] "gather and send the preceding
  values"
ifFalse: [ { {newField first}. {newField second} } ]
  "otherwise, set the two features as independant values and
  send the results"! !

!Athena methodsFor: 'gears' stamp: 'JohanBriglia 12/16/2016
  16:46'!
learn

"adds a new trace as calculated by the method calculateNewField
  "

mnesicFields add: (AthenaMnesicField newWith: self
  calculateNewField)! !

!Athena methodsFor: 'gears' stamp: 'JohanBriglia 12/16/2016
  16:52'!
gatherNewField

```

```

"optimized method to calculate gathered values
if the difference of nearby values is lesser than the fluence
value, they're gathered"

gatheredField := OrderedCollection new add: OrderedCollection
  new ; yourself.
gatheringInterval do: [ :x |
  gatheredField last add: (newField at: x).
  ((sensorimotorResonance at: x) - (sensorimotorResonance at: x
    + 1)) abs > fluence abs
  ifTrue: [ gatheredField add: OrderedCollection new ] ].

gatheredField last add: newField last.
^gatheredField! !

!Athena methodsFor: 'gears' stamp: 'JohanBriglia 12/16/2016
  16:37'!
inject: aCollection

"main gear of the model"

sensorimotorField := aCollection. "sets aCollection as the
  probe"
self calculateFluence ; "calculate the intensity L(theta|X)"
  calculateResonance. "calculate the echo P(theta|X)"
Athena isTesting "learns from this situation only if Athena is
  in integration mode"
  ifFalse: [ self learn ].

^fluence "returns the intensity value, useful for the recursive
  method named calculateFluence"! !

!Athena methodsFor: 'instance variables' stamp: 'JohanBriglia
  12/16/2016 16:10'!
archetypicalities

"similar with Minerva2 activations but in a more contextual way

```

```
"  
  
^archetypicalities ! !  
  
!Athena methodsFor: 'instance variables' stamp: 'JohanBriglia  
  12/16/2016 16:12'!  
similarities  
  
"similar to Minerva2 simiarities"  
  
^similarities! !  
  
!Athena methodsFor: 'instance variables' stamp: 'JohanBriglia  
  12/16/2016 16:10'!  
fluence  
  
"similar to Minerva2 intensity but in a more contextual way"  
  
^fluence ifNil: [ 0 ]! !  
  
!Athena methodsFor: 'instance variables' stamp: 'JohanBriglia  
  12/16/2016 16:10'!  
mnesicFields  
  
"list of traces within the model instance"  
  
^mnesicFields! !  
  
!Athena methodsFor: 'instance variables' stamp: 'JohanBriglia  
  12/16/2016 16:11'!  
sensorimotorResonance  
  
"response object similar to Minerva2's echo"  
  
^sensorimotorResonance ifNil: [ emptyField ]! !  
  
!Athena methodsFor: 'enumeration' stamp: 'JohanBriglia  
  12/16/2016 16:28'!
```

```
numberOfModalities
```

```
"number of features used by this instance of Athena while
  injecting a probe"
```

```
^mnesicFields first modalities size! !
```

```
!Athena methodsFor: 'enumeration' stamp: 'JohanBriglia
  12/21/2016 15:48'!
```

```
subAthenasNumber
```

```
"recursive method to count the total number of Athena instances
  (adds 1 to count itself)"
```

```
^(mnesicFields sum: [ :each | each subAthenasNumber ]) + 1! !
```

```
!Athena methodsFor: 'enumeration' stamp: 'JohanBriglia
  12/16/2016 16:28'!
```

```
numberOfTraces
```

```
"total number of traces in this instance of Athena"
```

```
^mnesicFields size! !
```

```
!Athena methodsFor: 'enumeration' stamp: 'JohanBriglia
  12/16/2016 16:29'!
```

```
subFieldsNumber
```

```
"recursive method to count the total number of Athena traces"
```

```
^mnesicFields sum: [ :each | each subFieldsNumber ]! !
```

```
!Athena methodsFor: 'initialization' stamp: 'JohanBriglia
  12/16/2016 16:27'!
```

```
testProbes: aBunchOfProbes
```

```
"test several probes and return results"
```

```

| results |

results := OrderedCollection new.

UIManager default "use a progress bar"
  displayProgress: 'Testing'
  from: 0 to: aBunchOfProbes size
  during: [ :bar |
    aBunchOfProbes doWithIndex: [:each :index |
      bar value: index.
      results add: (self testProbe: each) ]]. "send each probe to
      the testProbe method"

^results "and return the results"! !

!Athena methodsFor: 'initialization' stamp: 'JohanBriglia
  12/16/2016 16:13'!
initialize

"initialize the traces as an empty collection"

super initialize.
mnesicFields := OrderedCollection new! !

!Athena methodsFor: 'initialization' stamp: 'JohanBriglia
  12/16/2016 16:25'!
integrateProbe: aProbe

"inject a probe and return probe, echo and intensity"

Athena testing: false. "set Athena in Integration mode (it can
  learn)"
mnesicFields isEmpty "if there is no trace inside Athena"
  ifTrue: [ self startWith: aProbe ] "initialize the model with
    this first probe"
  ifFalse: [ self inject: aProbe ]. "otherwise inject the probe
    in the model"

^{ aProbe. self sensorimotorResonance. self fluence } "return

```

```

    results"! !

!Athena methodsFor: 'initialization' stamp: 'JohanBriglia
    12/16/2016 16:24'!
startWith: aCollection

"as the first probe, aCollection initialize the model"

mnesicFields add: ( AthenaMnesicField newWith: "add a new trace
    including "
    (aCollection collect: [ :each | { each } ])). "each feature
    as an independant number"

"The following presets recuring calculations for computer
    optimization "
emptyField := OrderedCollection new: aCollection size withAll:
    0.
gatheringInterval := (1 to: aCollection size - 1).

canProcreate := emptyField size > 2
    "this instance can't create sub-instances of Athena if there
    is no more than 2 features .
    there is no need to gather two features already gathered ,
    only computing time changes without this" ! !

!Athena methodsFor: 'initialization' stamp: 'JohanBriglia
    12/16/2016 16:27'!
integrateProbes: aBunchOfProbes

"integrate several probes and return results "

| results |

results := OrderedCollection new.

UIManager default    "use a progress bar"
    displayProgress: 'Integration '
    from: 0 to: aBunchOfProbes size
    during: [ :bar |

```

```

aBunchOfProbes doWithIndex: [:each :index |
  bar value: index.
  results add: (self integrateProbe: each)]]]. "send each
  probe to the integrateProbe method"

^results "and return the results"! !

!Athena methodsFor: 'initialization' stamp: 'JohanBriglia
  12/16/2016 16:26'!
testProbe: aProbe

"test a probe and return results including fluence and echo"

Athena testing: true. "set Athena in testing mode (without
  learning)"
mnesicFields isEmpty "if there is no trace inside Athena"
  ifTrue: [ self startWith: aProbe ] "initialize the model with
    this first probe"
  ifFalse: [ self inject: aProbe ]. "otherwise inject the probe
    in the model"

^{ aProbe. self sensorimotorResonance. self fluence } "return
  results"! !

"-----"!

Athena class
  instanceVariableNames: 'learn isTesting'!

!Athena class methodsFor: 'as yet unclassified' stamp: '
  JohanBriglia 12/16/2016 16:54'!
testing: aBoolean

"sets testing mode to true or false (integration mode)"

isTesting := aBoolean! !

!Athena class methodsFor: 'as yet unclassified' stamp: '
  JohanBriglia 12/16/2016 16:53'!

```

```
newWith: aCollection

"initialize the model with aCollection as the first trace"

^self new startWith: aCollection ; yourself! !

!Athena class methodsFor: 'as yet unclassified' stamp: '
  JohanBriglia 12/16/2016 16:53'!
isTesting

"returns true is Athena is in testing mode, false if in
  integration mode"

^isTesting! !

Object subclass: #AthenaMnesicField
  instanceVariableNames: 'modalities fluences resonance gathering
    stream'
  classVariableNames: ''
  poolDictionaries: ''
  category: 'ATHENA'!
!AthenaMnesicField commentStamp: 'JohanBriglia 12/16/2016 17:04'
  prior: 0!
An AthenaMnesicField is a trace within an instance of Athena. It
  can process trace features (AthenaModality instances) and
  sub-instances of Athena likewise.!

!AthenaMnesicField methodsFor: 'gears' stamp: 'JohanBriglia
  12/21/2016 15:37'!
inject: aCollection

"inject a probe in this trace in order to compare each other
  out"

stream := ReadStream on: aCollection. "reads the probe as a
  Stream to optimize computing"
```

```

fluences := gathering collectWithIndex: [ :each :index |
    (modalities at: index) inject: (stream next: each) ].
    "read the Stream according to features gathering and call the
    inject method accordingly
    i.e., with an Athena Instance or AthenaModality instance. The
    fluences collection is sent back"

resonance := modalities inject: #() into: [ :x :each | "collect
    features of contextual trace (echoes)"
    x, each sensorimotorResonance ].

^fluences average "send similarity value"! !

!AthenaMnesicField methodsFor: 'enumeration' stamp: '
    JohanBriglia 12/21/2016 15:39'!
subAthenasNumber

    "count instances of Athena within the trace"

^modalities sum: [ :each | each subAthenasNumber ]! !

!AthenaMnesicField methodsFor: 'enumeration' stamp: '
    JohanBriglia 12/21/2016 15:40'!
subFieldsNumber

    "count instances of traces within this trace (adds 1 to count
    itself)"

^(modalities sum: [ :each | each subFieldsNumber ]) + 1! !

!AthenaMnesicField methodsFor: 'initialization' stamp: '
    JohanBriglia 12/16/2016 17:08'!
setModalities: aCollection

    "set features and Athena instance(s) as this trace features (in
    a collection)"

```

```

modalities := aCollection! !

!AthenaMnesicField methodsFor: 'initialization' stamp: '
  JohanBriglia 12/16/2016 17:07'!
setGathering: aCollection

"set aCollection as the gathering values of a trace"

gathering := aCollection! !

!AthenaMnesicField methodsFor: 'instance variables' stamp: '
  JohanBriglia 12/16/2016 17:07'!
resonance

"contextual values of the trace (echo for an Athena instance,
  value for a feature)"

^resonance! !

!AthenaMnesicField methodsFor: 'instance variables' stamp: '
  JohanBriglia 12/16/2016 17:06'!
modalities

"features of a trace (AthenaMnesicField)"

^modalities! !

!AthenaMnesicField methodsFor: 'instance variables' stamp: '
  JohanBriglia 12/16/2016 17:06'!
gathering

"vector representing the gathering of features (modalities)"

^gathering! !

"-----"!

AthenaMnesicField class

```

```

instanceVariableNames: ''!

!AthenaMnesicField class methodsFor: 'as yet unclassified' stamp
  : 'JohanBriglia 3/25/2016 16:13'!
newWith: aMetaCollection

^self new setModalities:
  (aMetaCollection collect: [ :each |
    each size = 1
    ifTrue: [ AthenaModality newWith: each first ]
    ifFalse: [ Athena newWith: each ] ]) ;
  setGathering: (aMetaCollection collect: [ :each | each size ])
  ;
  yourself! !

Object subclass: #AthenaModality
  instanceVariableNames: 'x'
  classVariableNames: ''
  poolDictionaries: ''
  category: 'ATHENA'!
!AthenaModality commentStamp: 'JohanBriglia 12/21/2016 15:42'
  prior: 0!
An AthenaModality is an independant feature in the current trace
.
It uses the same methods as Athena to be used by the trace as
  any type of feature/model_instance!

!AthenaModality methodsFor: 'instance variables' stamp: '
  JohanBriglia 12/21/2016 15:45'!
sensorimotorResonance

"the contextual value of an independant feature is always the
  one that has been learned"

^{ x }! !

```

```
!AthenaModality methodsFor: 'enumeration' stamp: 'JohanBriglia
  12/21/2016 15:47'!
```

```
subAthenasNumber
```

```
"recursive method to count the total number of Athena instances
```

```
."
```

```
Ends here because an independant feature never contains Athena
  instances"
```

```
^0! !
```

```
!AthenaModality methodsFor: 'enumeration' stamp: 'JohanBriglia
  12/21/2016 15:48'!
```

```
subFieldsNumber
```

```
"recursive method to count the total number of traces.
```

```
Ends here, because an independant feature never contains traces
  "
```

```
^0! !
```

```
!AthenaModality methodsFor: 'gears' stamp: 'JohanBriglia
  12/21/2016 15:44'!
```

```
inject: y
```

```
"likelyhood estimator using infimum for interpretation
  optimization"
```

```
^(x + y first / 2) abs - (x - y first / 2) abs! !
```

```
!AthenaModality methodsFor: 'initialization' stamp: '
  JohanBriglia 12/21/2016 15:46'!
```

```
setValue: aValue
```

```
"sets the invariable feature value"
```

```
x := aValue! !
```

