

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LA PSYCHOTHÉRAPIE PSYCHODYNAMIQUE SELON LES
NEUROSCIENCES COGNITIVES : UNE INVESTIGATION DU POTENTIEL DU
PRINCIPE DE L'ÉNERGIE LIBRE POUR L'UNIFICATION DES SCIENCES DE
L'ESPRIT CONTEMPORAINES

MÉMOIRE

PRÉSENTÉ(E)

COMME EXIGENCE PARTIELLE

À LA MAÎTRISE EN PHILOSOPHIE

PAR

FLORIAN CHOQUET

MAI 2021

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.04-2020). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je souhaite remercier Pierre Poirier, qui est objectivement — il s'agit là d'une vérité métaphysique — le meilleur directeur de thèse qu'un étudiant puisse rêver d'avoir.

Je souhaite remercier ma maman, évidemment.

Je souhaite remercier mon meilleur ami Arnaud, mon fidèle compagnon de galère sans lequel je me sentirais bien seul sur cette terre.

Je souhaite remercier messieurs Serge Robert et Luc Faucher, d'avoir su tolérer les conditions extrêmes de la correction de mon mémoire de maîtrise.

À ensemble des professeurs et du personnel du département de philosophie de l'UQAM : merci à vous de m'avoir offert une formation et un soutien hors pair sans lequel je ne serais pas là où j'en suis aujourd'hui.

TABLE DES MATIÈRES

LISTE DES TABLEAUX.....	v
RÉSUMÉ	vi
INTRODUCTION	1
CHAPITRE I Explanandum : la psychothérapie psychodynamique	15
1.1 La psychanalyse freudienne.....	16
1.1.1 La théorie psychanalytique	16
1.1.2 La cure psychanalytique.....	23
1.2 La Neuropsychanalyse.....	27
1.3 La psychologie psychodynamique.....	36
1.4 La théorie de la mentalisation	43
1.5 La psychothérapie psychodynamique.....	49
CHAPITRE II Explanans : un modèle de l'esprit humain issu des sciences de l'esprit contemporaines	55
2.1 L'organisation fondamentale des systèmes vivants.....	56
2.1.1 L'autopoièse.....	57
2.1.2 Le principe de l'énergie libre	61
2.1.3 La neuroéthologie variationnelle.....	66
2.2 Application du principe de l'énergie libre au cerveau humain : une théorie computationnelle du fonctionnement neurologique.....	72
2.2.1 Les approches du traitement prédictif	72
2.2.2 L'allostasie, l'intéroception, et la conscience affective	75
2.2.3 Un modèle des émotions selon les approches du traitement prédictif	84
2.2.4 La neuropsychologie des grands réseaux neuronaux	92
2.3 L'échelle écologique et socioculturelle de la cognition humaine.....	100
2.3.1 Le « Skillfull Intentionality Framework ».....	100
2.3.2 Le modèle des affordances culturelles	102

2.4 Synthèse.....	104
CHAPITRE III Explication : un nouveau cadre théorique pour la psychothérapie psychodynamique issu des neurosciences cognitives contemporaines	108
3.1 L'espace d'échelle de la psychothérapie psychodynamique	109
3.1.1 L'Ego, le réseau du mode par défaut, et le Soi psychologique	109
3.1.2 L'émergence du Soi à travers le couplage structurel d'un individu avec l'écologie socioculturelle de l'espèce humaine	120
3.1.3 La mentalisation et le Soi psychologique selon les approches prédictives du cerveau humain	126
3.1.4 Un modèle de la communication humaine selon le principe de l'énergie libre et approches prédictives du cerveau humain	132
3.2 La psychologie psychodynamique reformulée	144
3.2.1 L'inconscient dynamique revisité	144
3.2.2 Le phénomène de l'attachement selon le principe de l'énergie libre et les approches prédictives du cerveau humain	168
3.2.3 Une reformulation du processus développemental d'acquisition de la capacité de mentalisation à travers les relations d'attachement.....	171
3.2.4 Les soins cognitifs et le prior de confiance épistémique.....	177
3.3 La psychothérapie psychodynamique reconceptualisée	185
3.3.1 Le niveau organisation des systèmes neurocognitifs humains adressé par la psychothérapie psychodynamique	186
3.3.2 La modification de l'organisation des systèmes neurocognitifs humains permise par la pratique de la psychothérapie psychodynamique	191
CONCLUSION.....	222
BIBLIOGRAPHIE.....	230

LISTE DES TABLEAUX

Tableau	Page
2.1 Les sens philosophiques et neurocognitifs des notions de catégorie et de concept.....	84

RÉSUMÉ

Ce mémoire constitue une tentative d'explication naturaliste et mécaniste de la pratique de la psychothérapie psychodynamique à l'aide des avancées théoriques récentes réalisées au sein des neurosciences cognitives contemporaines. En premier lieu, nous procédons à une revue des développements théoriques qui ont mené de la psychanalyse freudienne à la psychologie psychodynamique contemporaine, afin de nous constituer une compréhension initiale du cadre explicatif qui sous-tend la psychothérapie psychodynamique telle qu'elle est pratiquée aujourd'hui. Ensuite, nous présentons le cadre métathéorique du *Free Energy Principle* et son extension dans les approches prédictives du cerveau humain, ainsi qu'un ensemble de modèles théoriques intégratifs qui en découlent et qui ont été développés pour expliquer les dynamiques multiéchelles qui déterminent le fonctionnement cognitif humain. Ce cadre conceptuel et théorique va nous permettre de nous former un modèle général nous permettant de nous représenter l'ensemble des échelles d'organisation des systèmes neurocognitifs humains qui sont pertinentes pour comprendre les phénomènes en jeu au sein de la psychothérapie psychodynamique. Enfin, nous procédons à l'intégration de travaux récents afin d'élaborer une explication des mécanismes à travers lesquels l'activité de communication coopérative qui caractérise la psychothérapie psychodynamique peut mener à une modification significative du fonctionnement cognitif d'un individu.

Mots clés : psychothérapie psychodynamique, neurosciences cognitives, Free Energy Principle, approches prédictives

INTRODUCTION

Le travail réalisé au sein de ce mémoire de maîtrise en philosophie s'inscrit dans le domaine de général de la philosophie de l'esprit et des sciences cognitives¹, et a pour sujet principal la compréhension des phénomènes en jeu au sein de la pratique de ce que l'on appelle la « Psychothérapie ». L'association américaine de psychologie (APA) définit la psychothérapie comme l'application intentionnelle et informée de méthodes cliniques et de postures interpersonnelles dérivées de principes psychologiques établis dans le but d'assister la modification des comportements, des émotions, et de la personnalité des participants dans une direction qui leur semble désirable (Zeig et Munion, 1990). Dans une résolution portant sur l'efficacité générale de la psychothérapie (Campbell *et al.*, 2013) l'APA se base sur une revue extensive de la recherche empirique en psychologie pour faire les affirmations suivantes : la psychothérapie est efficace et ses effets sont constants à travers l'ensemble des troubles mentaux ; les effets de la psychothérapie durent plus longtemps et demandent moins de traitements additionnels que les traitements psychopharmacologiques. L'existence de la psychothérapie et son efficacité dans le traitement des troubles mentaux soulèvent plusieurs questions qui sont au cœur de la philosophie de l'esprit : le rapport entre nature et culture, entre gène et environnement ; le débat entre réductionnisme et

¹ Tout au long de ce mémoire, nous entendons "sciences cognitives" dans un sens large qui comprend, outre les disciplines fondatrices du domaine (la psychologie, la linguistique et l'Intelligence Artificielle), des disciplines qui s'y sont jointes plus récemment comme les neurosciences et l'anthropologie.

émergentisme, entre internalisme et externalisme de la cognition ; la nature des émotions et du « soi », etc.

Historiquement, le terme « psychothérapie » réfère à la pratique de la thérapie par la parole, élaborée par Sigmund Freud au sein de la psychanalyse, et dont les diverses variantes contemporaines sont réunies sous l'appellation de « psychothérapie psychodynamique ». Tout au long du 20^e siècle, de nouvelles formes de psychothérapies, correspondant à différentes approches théoriques en psychologie, furent développées. Les thérapies behaviorales appliquent les principes du conditionnement classique et opérant afin de modifier les comportements pathologiques des individus, et ce en se concentrant sur les aspects observables du fonctionnement psychologique humain. La Gestalt-thérapie adopte un point de vue existentialiste sur l'individu, en insistant sur l'idiosyncrasie de son expérience et sur sa capacité à être responsable. La thérapie humaniste adopte une approche positive et optimiste sur la nature humaine, cherchant à permettre au patient de s'autoactualiser à travers le développement de son potentiel. Grâce à l'avènement des sciences cognitives, de nouvelles formes de psychothérapies basées sur les connaissances empiriques nouvellement développées ont pu voir le jour. Ces pratiques sont réunies sous l'appellation générale de « thérapie cognitivo-comportementale » (TCC). La TCC peut être définie comme une modalité d'intervention psychosociale dont le but est d'améliorer la santé mentale des patients en se concentrant sur le traitement de leurs symptômes cognitifs et comportementaux actuels. La TCC inclut de nombreuses formes de thérapies spécifiques adressant des troubles mentaux et des symptômes définis et délimités par la recherche scientifique.

Cette approche systématique se prêtant particulièrement bien au processus de validation scientifique devenu la norme en psychologie, la TCC s'est très rapidement vu octroyé un statut privilégié au sein de la communauté scientifique internationale. Pourtant, toutes les autres formes de psychothérapies, incluant la psychothérapie

psychodynamique, continuent d'être pratiquées à travers le monde. Contrairement aux idées reçues, la littérature sur l'efficacité thérapeutique de la psychothérapie est loin d'avoir atteint un consensus sur l'efficacité comparative de la TCC et des autres formes de psychothérapies. Au contraire, il semblerait que toutes les formes de psychothérapie démontrent des niveaux relativement similaires d'efficacité (Castonguay *et al.*, 2006 ; Livesley, 2007 ; Norcross, 2019 ; Wampold, 1997 ; Wampold *et al.*, 2002) et que des facteurs communs comme la qualité de la relation entre le patient et le thérapeute soient les meilleurs prédicteurs de l'efficacité thérapeutique (Baldwin *et al.*, 2007 ; Flückiger *et al.*, 2011 ; Messer et Wampold, 2002 ; Wampold, 2015). De plus, la recherche sur l'efficacité thérapeutique des différentes psychothérapies est un champ d'études qui présente des problèmes méthodologiques importants (Wampold, 1997). Certains travaux ont montré que les résultats de plusieurs études sur l'efficacité thérapeutique pouvaient être expliqués par des biais causés par l'allégeance théorique des chercheurs qui les réalisaient (Dragioti *et al.*, 2015 ; Wilson *et al.*, 2011).

Aujourd'hui encore, il n'est pas rare que les approches psychothérapeutiques de descendance psychanalytique soient rejetées et discréditées par les chercheurs en psychologie (en particulier en Amérique du Nord), et ce en raison du caractère supposément « non-scientifique » de leur pratique. Pourtant, la recherche empirique sur l'efficacité thérapeutique de la psychothérapie ne démontre en aucun cas l'infériorité de la psychothérapie psychodynamique par rapport aux autres formes de psychothérapies (Fonagy, 2015 ; Shedler, 2010, 2018 ; Solms, 2018 ; Steinert *et al.*, 2017). Pour comprendre les raisons du discrédit qui est jeté sur la psychothérapie psychodynamique, il est important de saisir la différence entre les connaissances empiriques sur l'efficacité d'un traitement médical quel qu'il soit, et la validité théorique du système de croyances sur lequel il repose. En effet, si les critiques avancées à l'encontre de la psychologie psychodynamique sont suffisamment informées, ce n'est pas l'efficacité de sa pratique clinique (puisqu'elle est empiriquement prouvée) mais la validité du cadre théorique sur lequel elle repose, celui

de la psychanalyse, qui est remise en question. Plus précisément, il semble que ce soit la valeur explicative de la théorie psychanalytique qui soit critiquée.

Ici, et tout au long de ce mémoire, nous endossons une conception mécaniste de l'explication issue de la nouvelle philosophie mécaniste (Craver et Tabery, 2019 ; Glennan, 2017). Selon cette conception, l'explication en science consiste en l'explicitation des mécanismes qui sous-tendent un phénomène particulier. Dans ce contexte, tout mécanisme est mécanisme d'un phénomène. Le mécanisme d'un phénomène est ainsi défini comme un ensemble d'entités (ou de parties) dont les activités et les interactions sont organisées de telle manière qu'ils produisent le phénomène en question. Dans le cadre de la psychothérapie, le phénomène correspond à la modification positive du fonctionnement psychologique et de la santé mentale du patient à travers son interaction interpersonnelle avec le psychothérapeute. Le cadre théorique qui sous-tend une psychothérapie a une fonction explicative, il doit pouvoir expliciter le ou les mécanismes qui expliquent les modifications du fonctionnement psychologique du patient et qui ont lieu à travers son interaction avec le psychothérapeute.

En effet, bien que la psychologie psychodynamique propose bel et bien une explication qui implique l'explicitation d'un ensemble de mécanismes, ses détracteurs affirment que ces derniers ne répondent pas aux critères épistémiques des sciences de l'esprit contemporaines car : a) leur existence n'a pas été pas validée empiriquement par l'application de la méthode scientifique ; et b) il ne sont pas réductible à, ou ne trouvent pas de correspondance avec, les mécanismes proposés par les modèles théoriques des sciences cognitives. Les modèles théoriques des sciences cognitives sont issus de l'application de la méthode scientifique, ce qui leur procure un avantage épistémique qui justifie leur supériorité face à la psychologie psychodynamique. Dans ce contexte, et malgré son efficacité thérapeutique avérée, la psychothérapie psychodynamique se trouve rejetée par une grande partie de la communauté universitaire en psychologie

clinique qui va privilégier la TCC pour son cadre théorique validé scientifiquement. En conclusion, le rejet de la théorie psychanalytique par la communauté scientifique en psychologie semble reposer sur un conflit d'ordre épistémologique, qui résulterait de l'incapacité à mettre en relation les modèles théoriques de la psychologie psychodynamique et des sciences cognitives.

C'est pour répondre à cette lacune que les dernières décennies ont vu naître plusieurs tentatives d'actualiser la psychologie psychodynamique à l'aide des avancées réalisées dans les sciences de l'esprit contemporaines. Ces tentatives forment la base la plus solide pour le projet de développement d'un nouveau cadre théorique pour expliquer la psychothérapie psychodynamique. Pour cette raison, ce sont elles qui constitueront la base de notre mémoire, et ce sont elles que nous allons tenter d'expliquer, et non pas les modèles classiques de la psychologie psychodynamique. Parmi ces modèles contemporains, nous en avons sélectionné deux qui sortent du lot de par leur influence, leur compatibilité avec les neurosciences cognitives contemporaines, et leur respect des principes qui gouvernent la pratique clinique de la psychothérapie psychodynamique.

Le premier, la Neuropsychanalyse 2021-05-01 20:03:00, est un mouvement interdisciplinaire qui associe la théorie psychanalytique aux méthodes empiriques des neurosciences cognitives contemporaines afin de : a) identifier la nature et la structure des substrats neurologiques qui sous-tendent les phénomènes mentaux étudiés théorie psychanalytique ; b) élargir la portée des théories neuroscientifiques afin de pouvoir rendre compte de la dimension subjective des phénomènes mentaux ; et c) évaluer, réviser, et améliorer les concepts psychanalytiques à la lumière des connaissances actuelles en neurosciences cognitives (Fotopoulou *et al.*, 2012). Le projet de la neuropsychanalyse constitue le contexte général dans lequel va s'inscrire le travail réalisé dans le cadre de ce mémoire de maîtrise.

Le second modèle, la théorie de la mentalisation (Fonagy *et al.*, 2018), est selon nous la tentative contemporaine la plus aboutie pour expliquer les dynamiques du fonctionnement cognitif humain qui étayent et sous-tendent la pratique de la psychothérapie psychodynamique. La théorie de la mentalisation est une théorie constructiviste et sociale des émotions qui adopte la perspective de la théorie de l'esprit afin d'expliquer le développement émotionnel de l'individu. Selon la théorie de la mentalisation, les capacités cognitives qui permettent à un individu de se représenter, de comprendre, et de réguler ses états émotionnels sont acquises durant l'ontogénie à travers les relations intimes qu'il entretient avec ses donneurs de soins primaires (souvent les parents) et son environnement social proximal.

Il est de notre avis que les conflits qui émergent entre les différents modèles théoriques qui portent sur la pratique de la psychothérapie ne sont que des exemples particuliers d'un problème plus général qui touche l'ensemble des sciences de l'esprit. En effet, les sciences de l'esprit contemporaines forment un domaine scientifique éclaté en de multiples disciplines qui étudient des phénomènes connexes avec des méthodologies diverses : les neurosciences, les sciences cognitives, la psychologie (et toutes ses divisions internes), la neuropsychologie, l'éthologie, l'anthropologie cognitive, etc. Alors que le pluralisme méthodologique est une force, permettant d'investiguer les phénomènes étudiés sous différents angles, l'hétérogénéité et l'isolement des modèles théoriques développés dans ces différents champs d'études constituent un frein important qui ralentit les avancées scientifiques en empêchant la collaboration interdisciplinaire. Cette démultiplication disciplinaire, méthodologique, et théorique, est la conséquence directe de l'immense complexité du domaine de phénomènes qui caractérise ce champ d'études qu'est l'esprit.

Les défis épistémologiques posés par l'évolution des sciences de l'esprit ont mené de nombreux chercheurs et philosophes, incluant les partisans de la nouvelle philosophie mécaniste, à adopter la perspective suivante : la complexité des phénomènes

psychologiques (ou cognitifs), et le pluralisme disciplinaire qu'elle engendre, est due au fait qu'ils dépendent d'un grand nombre de mécanismes distribués à différentes échelles d'organisation des phénomènes naturels : allant des phénomènes physiques jusqu'aux phénomènes socioculturels en passant par les phénomènes biologiques. La nécessité de rendre compte des dynamiques multiéchelles du fonctionnement cognitif humain a mené à la popularisation du cadre conceptuel de la théorie générale des systèmes (Bertalanffy, 2003) qui confère au concept de système le statut d'unité épistémologique fondamentale.

La théorie des systèmes peut être définie comme l'étude interdisciplinaire de l'organisation abstraite des phénomènes, indépendamment de leur substance, type, ou des échelles spatiales et temporelles particulières dans lesquelles ils existent (Walonick, 1993). Elle étudie à la fois les principes communs à toutes les entités complexes, et les modèles (principalement mathématiques) qui peuvent être utilisés pour les décrire. Au sein de cette théorie, un système est défini comme un ensemble cohérent de parties interreliées et interdépendantes, qui peut être soit naturel, soit artificiel. Chaque système individuel est caractérisé par ses limites spatiales et temporelles, est entouré et influencé par son environnement, et est décrit par sa structure et son organisation qui s'expriment par l'intermédiaire de son fonctionnement (Heylighen et Joslyn, 1992)

L'un des courants théoriques principaux relevant de la théorie des systèmes est l'émergentisme, une position qui consiste à affirmer que, à partir de l'interaction non linéaire des composantes élémentaires d'un système, une toute nouvelle forme d'organisation peut émerger, laquelle possède des propriétés systémiques qualitativement distinctes de celles des composantes du système, même si elles émergent de leur interaction (les propriétés systémiques étant plus que la somme linéaire des propriétés des composantes du système). Dans le cadre de ce mémoire, nous suivons les partisans de la nouvelle philosophie mécaniste en adoptant une position émergentiste faible, composée d'une conception épistémique et mécaniste

(organisationnelle) de l'émergence. Cette forme d'émergentisme est compatible avec un réductionnisme mécaniste modéré, que nous endosserons par la même occasion. Cette conception de l'auto-organisation émergente des systèmes est très bien résumée par le concept d'organisation récursive qui, dans le cadre de la théorie générale des systèmes, peut être décrit par les principes suivants :

1. L'activité d'un système peut être décrite comme ayant différents niveaux d'organisation ;
2. L'organisation aux niveaux d'ordres supérieurs d'un système influence l'activité aux niveaux d'ordres inférieurs ;
3. Bien que les niveaux supérieurs d'organisation en viennent à définir (contraindre) l'activité aux niveaux inférieurs, ils ne peuvent pas violer les principes d'organisation de ces niveaux inférieurs ni excéder les limites d'action potentielles des éléments qui les composent (Connolly et van Deventer, 2017).

La théorie générale des systèmes nous offre un cadre conceptuel nous permettant de nous représenter la nature comme une immense hiérarchie composée de systèmes imbriqués les uns dans les autres. Chacun d'entre eux correspond à une échelle d'organisation à part entière qui, bien que présentant des propriétés uniques et irréductibles, ne peut pas s'extraire des lois qui régissent l'échelle d'organisation inférieure dont elle émerge. Ainsi, les systèmes socioculturels émergent de l'interaction entre les agents cognitifs humains, qui correspondent eux-mêmes à des systèmes (cognitifs) qui émergent des systèmes biologiques que sont les organismes qui composent l'espèce humaine, qui correspondent eux-mêmes à une forme particulière de système physique : les systèmes vivants.

Durant les dernières décennies, l'application de la théorie générale des systèmes aux sciences de l'esprit, en particulier au sein des neurosciences cognitives, a entraîné le développement de nouveaux modèles théoriques unificateurs tentant de rendre compte des dynamiques multiéchelles qui déterminent le fonctionnement psychologique humain. En effet, le champ d'études intégratif des neurosciences cognitives est lui-même issu de la volonté d'associer l'étude de la cognition humaine à l'investigation des mécanismes neurobiologiques sur lesquels elle repose. Dans ce mémoire, nous allons explorer les potentialités offertes par l'un de ces modèles : le modèle du « Free Energy Principle » (Friston, 2010) et son extension dans les approches prédictives du cerveau humain. Plus spécifiquement, nous allons appliquer ce modèle à la question de l'explication des phénomènes en jeu dans la pratique de la psychothérapie. Comme le sujet de la psychothérapie en général est beaucoup trop vaste pour être traité dans le cadre d'un mémoire de maîtrise, nous nous limiterons à la psychothérapie psychodynamique.

Ce choix est motivé par plusieurs raisons. Premièrement, en tant que descendante directe de la première forme de psychothérapie, la psychanalyse, la psychothérapie psychodynamique présente un grand nombre des caractéristiques fondamentales qui sont partagées par toutes les formes de psychothérapies. De plus, la psychothérapie psychodynamique, aux côtés, de la TCC, est la forme de psychothérapie la plus pratiquée. En la choisissant, nous espérons pouvoir contribuer à l'explication de la pratique de la psychothérapie en général.

Deuxièmement, nous pensons que l'incapacité de mettre en relation les modèles théoriques de la psychanalyse et des sciences cognitives était due à l'absence d'un cadre épistémologique permettant de rendre compte des dynamiques multiéchelles qui sous-tendent les phénomènes en jeu dans la pratique de la psychothérapie. Ce faisant, la question de l'explication de la psychothérapie psychodynamique constitue une étude de cas particulièrement intéressante pour démontrer le potentiel du FEP. En effet, nous

pensons que le modèle du FEP nous offre enfin les ressources théoriques suffisantes pour fournir une explication des phénomènes en jeu dans la psychothérapie psychodynamique qui soit compatible à la fois avec la psychologie psychodynamique et les sciences cognitives. Plus précisément, nous pensons que l'explication de ces phénomènes passe par l'explicitation des niveaux d'organisation auxquels ils correspondent, des principes auto-organisation qui les définissent, ainsi que de leur place dans l'échafaudage hiérarchique des systèmes cognitifs humains.

Troisièmement, il est de notre avis que le sous-ensemble spécifique des phénomènes psychologiques sur lequel la psychologie et la psychothérapie psychodynamique se concentrent est d'une importance capitale si nous voulons développer un modèle complet du fonctionnement cognitif humain. Nous sommes conscients du fait que la psychanalyse s'est longtemps extraite du processus de validation scientifique, et d'accord avec l'idée que cela est inacceptable. Cependant, la métaphore qui éclaire notre démarche consiste à dire « qu'il ne faut pas jeter le bébé avec l'eau du bain ». Nous sommes convaincus que l'efficacité de la psychothérapie psychodynamique est la conséquence du fait qu'elle agit sur certains aspects fondamentaux du fonctionnement cognitif humain qui peuvent être minimalement réduits à deux catégories non exclusives : a) la dimension affective et émotionnelle de la cognition humaine ; et b) les fonctions et activités cognitives qui sont dépendantes de, et/ou portent sur, l'environnement social de l'espèce humaine. Selon la théorie générale des systèmes et la nouvelle philosophie mécaniste, les mécanismes qui sous-tendent ces aspects de la cognition humaine devraient être distribués à différentes échelles d'organisation des systèmes cognitifs humains.

Notre thèse

La thèse principale qui sera défendue dans notre mémoire est la suivante :

- Il est possible de formuler une explication des phénomènes en jeu au sein de la psychothérapie psychodynamique qui soit compatible avec les intuitions fondamentales de la psychanalyse freudienne et de la psychologie psychodynamique, tout en étant réductible à des théories scientifiques actuelles, que ce soit en psychologie, en neurosciences cognitives, en anthropologie cognitive, etc.

Nos problématiques

Pour défendre cette thèse, nous allons tenter d'étayer la formulation d'une réponse à plusieurs questions qui sont explicitées dans les problématiques suivantes :

- À quel(s) niveau(x) d'organisation spécifique(s) des systèmes neurocognitifs humains correspondent les phénomènes impliqués dans la pratique de la psychothérapie psychodynamique ? Quels sont les mécanismes ou les processus qui définissent l'organisation de ce(s) niveau(x) ?
- Quels sont les mécanismes à travers lesquels la psychothérapie psychodynamique amène une amélioration du fonctionnement cognitif et de la santé mentale d'un individu ? Comment ces mécanismes affectent-ils l'organisation du ou des niveau(x) des systèmes neurocognitifs humains auxquels ils correspondent ?

Notre plan

Afin de formuler des pistes de réponses aux questions mentionnées ci-dessus, notre mémoire sera composé de trois parties :

1. Dans la première partie intitulée *explanandum*, nous allons procéder à une introduction à la théorie psychanalytique et à la psychologie psychodynamique, ainsi qu'à un exposé des modèles de la Neuropsychanalyse et de la théorie de la

mentalisation afin de nous constituer une compréhension initiale des phénomènes sur lesquels reposent la pratique de la psychothérapie psychodynamique, ainsi qu'une conception provisoire des mécanismes dont dépendent son fonctionnement et son efficacité.

2. Dans la deuxième partie intitulée *explanans*, nous allons présenter le cadre métathéorique du Free Energy Principle ainsi que les modèles théoriques intégratifs qui en découlent et qui ont été développés pour expliquer les dynamiques multiéchelles du fonctionnement neurocognitif humain. Ce cadre conceptuel et théorique va nous permettre de nous former un modèle général permettant de nous représenter l'ensemble des échelles d'organisation des systèmes neurocognitifs humains qui sont pertinentes pour comprendre les phénomènes en jeu au sein de la psychothérapie psychodynamique.

3. Dans la troisième partie intitulée *explication*, nous allons intégrer des travaux récents réalisés dans le cadre des approches théoriques avancées dans notre *explanans* afin d'étayer la formulation d'hypothèses visant à identifier : a) le ou les niveaux d'auto-organisation spécifiques des systèmes cognitifs humains impliqués dans la pratique de la psychothérapie psychodynamique ; et b) les mécanismes à travers lesquels la psychothérapie psychodynamique participe à la modification de l'organisation de ce(s) niveau(x) permettant l'amélioration du fonctionnement cognitif et de la santé mentale des individus.

Notre objectif

En qualité de travail philosophique l'objectif de notre mémoire de maîtrise ne sera pas de déterminer l'efficacité thérapeutique de la psychothérapie², ce qui est une tâche qui relève de la recherche empirique. La question de savoir si, à partir des connaissances actuelles, nous pouvons affirmer ou non que la psychothérapie est une méthode d'intervention efficace pour le traitement des troubles mentaux, dépasse nos compétences ainsi que la portée d'un mémoire de maîtrise en philosophie. Une revue significative de la littérature au sujet de l'efficacité thérapeutique de la psychothérapie constituerait un travail trop important pour être incluse dans notre mémoire, qui présente déjà un niveau de complexité important. De plus, un travail philosophique sur les enjeux épistémologiques de ce domaine de recherche pourrait constituer le sujet de mémoire de maîtrise (ou de doctorat) à lui seul. C'est pourquoi nous avons choisis d'adopter l'efficacité thérapeutique de la psychothérapie comme prémisse à notre mémoire de maîtrise, et nous acceptons pleinement le caractère hypothétique que cela pourrait lui octroyer. Pour tous doutes quant au bien fondé de l'adoption de cette prémisse, nous dirigeons le lecteur vers la déclaration de l'APA mentionnée précédemment (Campbell *et al.*, 2013), ainsi que vers les nombreuses revues et méta-analyses préexistantes (Messer et Wampold, 2002 ; Munder *et al.*, 2018 ; Wampold, 1997, 2007 ; Wampold *et al.*, 1997 ; Wampold et Imel, 2015).

L'objectif de notre mémoire ne sera pas non plus de défendre la validité théorique de la psychanalyse freudienne ou de la psychologie psychodynamique de la

² Nous n'aurons pas non plus pour but de déterminer l'efficacité thérapeutique relative de la psychothérapie psychodynamique par rapport à d'autres formes de psychothérapie. Tel que mentionné précédemment, la recherche semble suggérer que la psychothérapie psychodynamique est au moins aussi efficace, si ce n'est plus dans certains cas, que d'autres formes de psychothérapie (cf. page blabla).

deuxième moitié du 20^e siècle en tant que telles. Tel que mentionné précédemment, il existe aujourd'hui des modèles théoriques intégratifs contemporains (cf. la Neuropsychanalyse et la théorie de la mentalisation) qui, tout en conservant certaines intuitions et hypothèses importantes de la psychanalyse et de la psychologie psychodynamique, présentent une bien meilleure validité théorique et une beaucoup plus grande compatibilité avec les neurosciences cognitives (et les autres sciences de l'esprit) contemporaines. Ce sont ces intuitions et ces hypothèses que nous souhaitons conserver afin de formuler un nouveau cadre théorique permettant d'expliquer les phénomènes en jeu dans la psychothérapie psychodynamique.

Notre objectif sera, à partir de ces théories intégratives, et de leur reformulation à l'aide d'un ensemble modèles issus d'avancées théoriques récentes au sein des neurosciences cognitives contemporaines (cf. *Le Free Energy Principle* et les approches prédictives du cerveau humain, entre autres), de contribuer au développement d'un nouveau cadre explicatif des phénomènes en jeu au sein de la psychothérapie psychodynamique. Plus précisément, notre but sera de participer à l'élaboration d'un modèle théorique opérationnel selon les critères méthodologiques des neurosciences cognitives contemporaines, afin d'étayer la formulation d'hypothèses afin de supporter la recherche empirique sur les mécanismes qui soutendraient l'efficacité de la psychothérapie psychodynamique.

CHAPITRE I

EXPLANANDUM : LA PSYCHOTHÉRAPIE PSYCHODYNAMIQUE

Dans ce premier chapitre de notre mémoire de maîtrise, nous allons procéder à un exposé introductif de l'ensemble des modèles théoriques qui constituent le cadre explicatif sur lequel est basée la psychothérapie psychodynamique telle qu'elle se pratique aujourd'hui. Par la même occasion, nous allons tenter d'expliciter le processus d'évolution théorique qui a mené de la théorie psychanalytique freudienne à la psychologie psychodynamique, puis au développement des deux modèles théoriques contemporains qui nous intéressent, à savoir la Neuropsychanalyse et la théorie de la mentalisation. Si nous avons choisis ces deux modèles théoriques, c'est parce qu'ils conservent les intuitions et les hypothèses fondamentales qui sous-tendent la pratique de la psychothérapie psychodynamique tout en les reformulant dans les termes des sciences de l'esprit contemporaines. Ainsi, ce premier chapitre aura deux objectifs : 1) expliciter les hypothèses et les intuitions fondamentales qui sous-tendent la pratique de la psychothérapie psychodynamique, et que nous devons étayer si nous voulons en fournir une explication satisfaisante ; et 2) nous constituer un cadre explicatif initial pour la psychothérapie psychodynamique qui présentera la meilleure compatibilité possible avec une reformulation ultérieure (au chapitre 3) à l'aide des avancées théoriques récentes au sein des neurosciences cognitives contemporaines (présentées au chapitre 2).

La description la plus simple que l'on peut donner de la psychothérapie psychodynamique est qu'il s'agit d'une méthode d'interaction d'ordre psychologique

utilisée pour soulager la souffrance émotionnelle d'une personne. Nous en fournirons une définition provisoire à la fin de cette section. Elle dérive de la première forme moderne de psychothérapie, la psychanalyse. Dans sa pratique en tant que neurologue clinicien, Sigmund Freud, fondateur de la psychanalyse, fût confronté à des pathologies (ex. Hystérie) dont la symptomatologie, la paralysie de certains membres par exemple, ne présentait pas de cohérence avec l'organisation nerveuse effective dans les parties du corps qui étaient touchées. C'est ainsi qu'il fut contraint, par la force des choses, de s'adonner à contrecœur à une théorisation qu'il qualifiait de « purement psychologique » (Freud, 1886). Freud développa la psychanalyse à la fois comme une méthode de recherche et comme une technique d'intervention, dans le but de comprendre et de traiter ces pathologies mentales (psychopathologies) dont il situait l'étiologie dans un dysfonctionnement de l'appareil psychique de ses patients. C'est ainsi qu'est née la première forme de psychothérapie, la thérapie par la parole, qui sera vite baptisée « cure psychanalytique » ou « psychanalyse » tout simplement. Cette pratique thérapeutique, ainsi que toutes les formes de psychothérapies développées subséquentement, repose sur une hypothèse fondamentale selon laquelle il est possible de comprendre et de traiter certains troubles mentaux à travers la communication verbale et l'observation du comportement de l'individu qui en est la victime.

1.1 La psychanalyse freudienne

1.1.1 La théorie psychanalytique

En 1895, cinq ans avant la parution de *L'interprétation des rêves* (Freud, 2010), ouvrage phare qui marque la naissance de la théorie psychanalytique, Sigmund Freud entreprit d'élaborer une théorie mécaniste de l'esprit. Inspiré par la récente théorie du

neurone de Ramón y Cajal et Waldeyer-Harz, le jeune neurologue qu'il était développa dans son *Projet pour une psychologie scientifique* (Freud, 1954) une théorie spéculative dans laquelle il expliquait le fonctionnement de l'esprit en matière d'échanges énergétiques entre des neurones aux fonctions particulières. Malheureusement, confronté aux limites de la biopsychologie de son époque, il fut contraint de renier ce travail qui ne fut publié qu'à titre posthume. Les réflexions ayant vu le jour au sein de cette tentative précoce vont néanmoins constituer la base du modèle métathéorique spéculatif sur lequel va reposer le développement de la psychanalyse : la métapsychologie.

Homme de son époque et médecin de formation, Freud était un matérialiste endurci et adoptait une forme de monisme naturaliste qui le poussa, entre autres, à rejeter la distinction diltheyenne entre les sciences de la nature et les sciences de l'esprit : il pensait qu'il n'y avait de science que des sciences de la nature (Assoun, 1990). Il n'est donc pas surprenant que les principes fondamentaux qui composent sa métapsychologie proviennent des modèles des sciences de la nature les plus influents de son époque. Le premier de ces principes dérive des travaux d'Helmoltz, et plus précisément de son application du principe physique de conservation de l'énergie, selon lequel l'énergie totale d'un système isolé est invariante au cours du temps, au domaine de la physiologie (Assoun, 1990). En effet, en partant de l'hypothèse selon laquelle un système nerveux humain est un système physique comme un autre, Freud formule le *principe de constance* qui affirme que la fonction première du système nerveux central est de débarrasser l'organisme de l'excès de stimulation et d'excitation afin de le maintenir à un niveau d'énergie constant (Freud, 2001). Sur la base de ce principe, Freud formule l'hypothèse suivante : chaque expérience est accompagnée d'un « quota d'affect », la conséquence psychique de l'excitation du système nerveux, et le rôle de l'appareil psychique (l'esprit) est de permettre la décharge de cette quantité d'affects.

L'hypothèse selon laquelle toute expérience psychique est accompagnée d'une quantité d'excitation (et d'une charge affective) pose la question de savoir d'où provient cette dernière. C'est pour répondre à cette question que Freud développa sa fameuse *théorie des pulsions*, dans laquelle il identifie les principales sources d'excitation qui parviennent à l'appareil psychique à un ensemble de motivations instinctuelles, innées, qui sont ancrées dans la structure biologique de l'individu et ont été façonnées par le processus de sélection naturelle afin de favoriser la survie et le succès reproducteur de l'organisme. Ces forces motivationnelles ont pour fonction d'éliciter les comportements permettant la satisfaction des besoins biologiques fondamentaux de l'organisme : faim, soif, température, sexualité, etc. Freud utilise le concept de *pulsion*, qu'il situe à la frontière des domaines de la biologie et de la psychologie, pour référer à l'activité psychique issue de ces forces motivationnelles. La théorie des pulsions permet à Freud d'affiner la fonction qu'il attribue à l'appareil psychique. En effet, si la fonction de l'esprit est de décharger l'excitation qui lui parvient, et que cette excitation provient des pulsions, alors l'une des tâches principales qu'il doit réaliser est de permettre la satisfaction des pulsions (Eagle, 2011).

Ainsi, une pulsion est une source d'excitation endogène à l'organisme qui exerce une pression sur le système nerveux (et l'appareil psychique) afin de permettre la satisfaction du besoin biologique qu'elle représente. Cette satisfaction doit s'opérer à travers l'interaction avec un *objet*, une entité externe dont les propriétés lui permettent de remplir le besoin en question. Contrairement à un *instinct*, qui correspondrait à un programme cognitivo-comportemental héréditaire et relativement fixe, la pression exercée par la pulsion n'est pas associée à des comportements ou à des objets prédéterminés. Cette nature relativement indéterminée de la pression motivationnelle exercée par les pulsions a deux conséquences : a) les comportements et les objets spécifiques choisis par un individu pour satisfaire ses pulsions sont le fruit d'un processus développemental dépendant de l'expérience ; et b) les pulsions peuvent être

satisfaites à travers une multitude de comportements et d'objets différents, pour autant qu'ils remplissent efficacement le besoin biologique auxquelles elles correspondent.

La quantité d'affects issue des pulsions cherche naturellement à s'exprimer dans l'activité de l'individu, et peut le faire de plusieurs manières. La manière la plus directe de décharger cette excitation est de le faire à travers l'action, un comportement visant à trouver et à interagir avec un objet physique concret permettant la satisfaction du besoin qu'elle manifeste, et c'est ce que font l'ensemble des animaux non humains. Freud affirme que pour un être humain, l'affect est le plus souvent déchargé à travers l'expérience consciente, ce qui inclut principalement l'utilisation du langage afin de le nommer et de parler du sentiment qu'il procure. En effet, Freud définit la capacité de pensée spécifique à l'espèce humaine comme une forme d'inhibition de l'action, qui serait remplacée par une activité mentale dans laquelle des représentations langagières (ou autres) vont venir contenir et permettre l'expression de l'affect (Macmillan, 2000). Ici, la pensée est conçue comme une forme d'hallucination permettant à l'individu de se représenter mentalement les objets qui lui permettraient potentiellement de satisfaire ses pulsions. Cette capacité serait exploitée afin de permettre de délayer la satisfaction des pulsions et de pouvoir planifier et mettre en place des actions plus complexes et adaptatives.

De plus, la quantité d'affects qui accompagne chaque expérience va pouvoir être diminuée en connectant les idées qui lui sont liées avec d'autres contenus mentaux élaborés à partir de la mémoire de l'individu. Dans ce processus appelé *l'association*, chaque représentation va générer une expérience qui lui est propre, ce qui va venir influencer l'intensité affective reliée à l'expérience initiale. Freud attribuait un second rôle à l'association, qui servirait une fonction de rectification cognitive, à travers laquelle le contenu représentationnel des idées associées à certaines expériences peut être modifié pour former une conception plus adéquate ou plus adaptative des expériences vécues. Le réseau formé par l'ensemble de ces contenus mentaux ainsi que

les liens qui les unissent est appelé « le grand complexe des associations », considéré par Freud comme constituant la structure fondamentale de la personnalité d'un individu (Eagle, 2011).

Lorsqu'une expérience est accompagnée d'un quota d'affect trop important, et particulièrement lorsque, pour une raison ou une autre, l'expression de l'affect est inhibée, cela rend la décharge de l'excitation par la représentation et l'association des idées plus difficile. Lorsqu'aucune de ces tâches ne peut s'effectuer correctement, comme c'est le cas dans l'hystérie, la charge affective est maintenue dans un état inerte, ce qui mène selon Freud à la formation de symptômes, qui résultent de la nécessité de diminuer l'excitation (principe de constance) accumulée dans l'appareil psychique d'une manière ou d'une autre. Dans le cas de l'hystérie par exemple, il s'agirait d'une part de la conversion de l'affect en symptômes somatiques comme une tension ou une paralysie musculaire, et d'autre part de l'isolation des idées problématiques du reste des contenus mentaux qui composent la psyché de l'individu. Tirée de la psychiatrie française de son temps, l'idée que l'isolation de contenus mentaux aurait des effets pathogènes qui seraient à l'origine de certaines psychopathologies est l'une des hypothèses fondamentales de la psychanalyse (Eagle, 2011). L'apport original de Freud réside dans la formulation du mécanisme de « refoulement », qui consiste à affirmer que l'isolement de contenus mentaux problématiques repose sur un acte mental qui *les exclut systématiquement de l'expérience consciente du sujet*, et ce afin d'éviter d'éliciter la charge affective qui leur est lié, et de maintenir l'excitation du système nerveux à un niveau acceptable (Laplanche et Pontalis, 2009). Le refoulement est ici conçu comme un mécanisme normal et adaptatif du fonctionnement psychique, dont l'excès ou le déficit peuvent tous les deux mener à des conséquences pathologiques.

Selon Freud, le refoulement mène à la création de groupes psychiques autonomes, qui d'après la théorie des pulsions sont composés non seulement d'idées mais aussi de

motivations, de buts ou de désirs, dont le contenu est isolé et différent de celui de la personnalité centrale de l'individu telle que reflétée dans son expérience consciente. Ces groupes psychiques sont conçus comme restant actifs et fonctionneraient parallèlement à l'activité consciente de l'individu. La théorie psychanalytique affirme que les symptômes psychopathologiques émergent lorsque l'activité et les objectifs de ces groupes psychiques refoulés entrent en contradiction avec ceux de la personnalité consciente de l'individu, une situation à laquelle on réfère à travers le terme de *conflit intrapsychique* (Freud *et al.*, 2016). Ici, les effets pathogènes du refoulement résultent de l'impossibilité d'associer, et donc d'intégrer le contenu mental de ces groupes psychiques isolés au reste du grand complexe des associations, ce qui mènerait à une scission et à une fragilisation de la personnalité de l'individu.

D'affirmer que les désirs restent actifs malgré le refoulement et l'isolation associative, c'est-à-dire, malgré le fait qu'ils ne sont pas directement expérimentés consciemment, n'est rien d'autre que de postuler l'existence de processus mentaux inconscients. Cela correspond à la thèse pour laquelle Freud est le plus connu, et qui consiste à affirmer qu'il existe des états mentaux inconscients, et que ces derniers constituent la majeure partie des états mentaux qui composent notre vie mentale. Selon la célèbre métaphore freudienne, la conscience, une acquisition tardive (phylogénétiquement) propre à l'être humain, n'occupe que le sommet de cet iceberg qu'est l'esprit. Cette thèse renvoie à ce que l'on appelle aujourd'hui l'inconscient descriptif, et elle a depuis longtemps été adoptée par la psychologie cognitive (Power et Brewin, 1991). Cependant, la conception freudienne de l'inconscient ne se limite pas à cette thèse, mais s'accompagne d'un modèle théorique sur la formation et la constitution de la partie inconsciente de l'esprit, et c'est ce dernier qui est la marque distinctive de la théorie psychanalytique. Appelée *l'inconscient dynamique*, cette partie de l'esprit est conçue comme étant composée de groupes psychiques composés de contenus mentaux (souvenirs, représentations, désirs, etc.) organisés autour des pulsions, des motivations

instinctuelles qui ont été isolées de l'expérience consciente à travers le mécanisme du refoulement (Kihlstrom, 2015).

Dans la psychanalyse freudienne, la structure de l'appareil psychique est représentée à l'aide d'une métaphore spatiale (dite « topique ») mettant en scène différentes instances psychiques aux fonctions psychologiques différentes, et dont la position dans l'esprit est déterminée par le caractère plus ou moins conscient ou inconscient des activités mentales qu'elles prennent en charge (Laplanche et Pontalis, 2009). Composé des pulsions refoulées, *l'Inconscient* (aussi appelé « Ça ») est théorisé comme la partie de l'appareil psychique qui détermine la plus grande partie du comportement humain. Il est rendu manifeste à travers les modes de cognition primitifs que l'on peut observer dans des états mentaux non ordinaires comme le rêve ou la psychose, et serait sous-tendu par les « processus primaires », un type d'activité cérébrale spécifique dans lequel l'échange d'énergie entre les neurones serait relativement libre (Carhart-Harris et Friston, 2010). Selon Freud, ces modes de cognition répondent au *principe de plaisir*, qui stipule qu'ils visent la décharge de l'excitation et la satisfaction des pulsions de la manière la plus directe et immédiate possible.

Freud reconnaissait dans ces états mentaux non ordinaires la perte de certaines fonctions, qui sont normalement présentes dans la cognition éveillée. Carhart-Harris et Friston (2010) affirment que Freud attribuait ces fonctions à une organisation centrale, *l'Ego* (ou « Moi »), sous-tendu par les « processus secondaires », un type d'activité neuronale ayant pour but la conversion de « l'énergie libre » en « énergie liée ». Selon Freud, l'Ego répond au *principe de réalité*, ce qui veut dire qu'il a pour fonction de réguler l'activité cognitive et comportementale de l'organisme afin qu'elle se traduise dans des comportements adaptatifs permettant de remplir efficacement ses besoins biologiques tout en répondant aux exigences de l'environnement physique et social de l'individu. Freud considère l'Ego comme le siège de la cognition consciente, des habiletés langagières, et de la raison, ainsi que comme le cœur de la personnalité de

l'individu. C'est l'Ego qui performe le refoulement, ainsi qu'une constellation d'autres mécanismes de défense (voir Vaillant, 1992) qui visent à empêcher que les motivations pulsionnelles ne se traduisent en une activité psychique ou des comportements inadaptés qui résulteraient, à terme, en une excitation excessive du système nerveux. Il est important de noter que, bien qu'étant le siège de l'expérience consciente, une partie importante des fonctions de l'Ego se déroulent inconsciemment, ce qui est le cas des mécanismes de défense (incluant le refoulement) mentionnés ci-dessus.

Dans la psychanalyse freudienne, les conflits intrapsychiques qui sont conçus comme étant à l'origine des psychopathologies comme l'hystérie ou la névrose obsessionnelle opposent les pulsions inconscientes aux défenses de l'Ego. Les symptômes qui caractérisent ces pathologies mentales sont conçus comme une formation de compromis entre ces deux instances psychiques, les fonctions de l'Ego contraignant la charge affective provenant des pulsions à s'exprimer par des voies détournées (ex. tension musculaire, idées obsessionnelles, tics comportementaux, anxiété, etc.). Freud affirma à de nombreuses reprises que le domaine d'application de la cure psychanalytique se limite exclusivement aux pathologies qui seraient issues d'un tel conflit psychique interne, et auxquelles il référerait à travers la catégorie nosologique des *névroses* (Freud, 1995 ; Waelder, 1960). Les difficultés d'expression de l'affect qui caractérisent l'ensemble de ces pathologies résultent en une souffrance émotionnelle importante et handicapante, et le but principal de la cure psychanalytique est de soulager cette dernière en résolvant les conflits intrapsychiques qui en sont à l'origine.

1.1.2 La cure psychanalytique

L'ensemble des principes et des concepts exposés jusqu'à maintenant sont suffisants pour nous permettre de fournir une description minimale de la cure psychanalytique. Cette méthode repose sur plusieurs principes et techniques qui ont tous pour but de

lever la barrière du refoulement afin de permettre aux contenus inconscients d'accéder à l'expérience consciente. Le passage de l'inconscient à la conscience aurait deux fonctions principales. La première serait de permettre à la charge affective qui leur est associée de s'exprimer, un processus appelé *abréaction* (ou catharsis). Au début de la psychanalyse l'abréaction était considérée comme l'objectif premier du traitement, un but qui reposait sur la croyance selon laquelle les excès d'excitations étaient principalement causés par des souvenirs reliés à des événements traumatiques (Eagle, 2011). C'est la technique l'hypnose qui était privilégiée pour aider les patients à accéder aux souvenirs traumatiques en question. Les traumatismes correspondant à des stimuli externes ponctuels, la charge affective qui leur était associée ne demandait à être déchargée qu'une seule fois. Cependant, une fois la théorie des pulsions développée, l'abréaction fut mise de côté au profit de la fonction de *rectification associative*, permettant la modification de l'intensité affective et du contenu représentationnel des contenus mentaux. En effet, les pulsions étant des sources endogènes et intarissables d'excitation, l'abréaction ne pouvait plus constituer qu'une solution partielle et passagère à l'excès d'excitation subit par l'appareil psychique.

Pour Freud, le passage de l'inconscient à la conscience était aussi synonyme de la possibilité pour l'excitation et les motivations instinctuelles provenant des pulsions, auparavant exprimées à travers les processus primaires, de trouver leur expression dans l'activité de l'Ego et des processus secondaires (Freud, 1995). Au-delà de permettre l'unification de la personnalité en autorisant les contenus inconscients à entrer dans le grand complexe des associations, et ainsi se débarrasser des demandes importantes du travail de refoulement constant qui les maintenaient isolées, l'expression des pulsions à travers les processus secondaires aurait pour but de permettre leur conversion en un comportement et une activité psychique efficaces et adaptatifs.

Selon Eagle (2011), c'est pour favoriser l'accomplissement de la fonction de rectification associative que fut développée la plus fondamentale des techniques

utilisées dans la cure psychanalytique, *l'association libre*, qui consiste à demander au patient de parler de tout ce qui lui vient à l'esprit, en s'abstenant du mieux qu'il puisse d'appliquer une quelconque censure morale ou logique à son propos (Laplanche et Pontalis, 2009). Basée sur le postulat d'un déterminisme psychique, l'association libre repose sur l'hypothèse que de lever le maximum de contraintes imposées à la pensée et à l'expression langagière du patient devrait autoriser la manifestation d'un échantillon plus exhaustif de l'ensemble des contenus mentaux qui présentent une charge affective importante pour lui, et particulièrement ceux qui dérivent des groupes psychiques inconscients issus du refoulement.

Bien évidemment, la technique de l'association libre ne pourra pas complètement contourner les défenses de l'Ego, et les pulsions seront tout de même contraintes de se manifester de manière détournée dans l'activité psychique du patient. Si la méthode thérapeutique élaborée par Freud s'appelle la psychanalyse, c'est qu'elle consiste essentiellement en l'analyse, principalement par le thérapeute, des contenus mentaux du patient afin d'identifier leur source et leur signification. Ce processus, *l'interprétation*, est la deuxième technique fondamentale de la cure psychanalytique. Les manifestations des défenses de l'Ego durant le processus thérapeutique sont appelées des *résistances*, et leur interprétation est conçue comme l'un des objectifs centraux de la cure psychanalytique, puisqu'elle permet la prise de conscience des processus qui sont à l'origine de l'isolation pathogène des contenus mentaux (Eagle, 2011).

Au fur et à mesure de son expérience clinique, Freud se rendit compte que les simples propos du patient n'étaient pas la seule et unique source d'accès aux contenus inconscients de son esprit. En effet, il fut forcé de constater que, au cours du traitement, le patient expérimente envers la personne du thérapeute des sentiments, des désirs, des attitudes, et des défenses qui ne sont pas propres à leur relation, mais qui trouvent leur origine dans les relations passées du patient avec des figures significatives, le plus

souvent les figures parentales, qui ont eu lieu durant l'enfance, et qui sont déplacées inconsciemment sur la figure du thérapeute (Greenson, 1978). Freud appela ce phénomène le *transfert*, et l'interprétation des manifestations comportementales et psychiques qui en résultent devint l'objectif central de la cure psychanalytique. Le transfert n'est pas limité au cadre de l'analyse, mais se déroule constamment dans l'ensemble des relations émotionnellement significatives entretenues par l'individu, et Freud pensait qu'il reflétait directement le cœur des conflits internes à l'origine des névroses. Puisque les réactions transférentielles ont lieu dans l'ici et maintenant, Freud considérait que la prise de conscience résultant de leur interprétation devrait s'imposer avec une immédiateté, une tangibilité et une vivacité émotionnelle supérieure à ce que pourrait procurer l'analyse de n'importe quel souvenir ou représentation mentale, ce qui aurait selon lui plus de chance de résulter en un changement durable de l'organisation de l'appareil psychique du patient.

Le cadre clinique de la psychanalyse est pensé pour favoriser l'application et le succès de l'ensemble des techniques énoncées jusqu'à présent. Le thérapeute adopte une posture psychologique particulière, qui consiste à restreindre son investissement émotionnel dans la relation avec le patient afin de : a) influencer le moins possible les contenus psychiques qui émergent durant la thérapie, et ce de manière ce qu'ils soient le plus représentatifs des enjeux du patient ; b) ne pas céder aux demandes du patient visant à la satisfaction de ses pulsions, et ce de manière à le pousser à la place à prendre conscience de ces motivations instinctuelles afin qu'elle puissent être intégrées au reste de sa personnalité ; et c) ne pas reproduire les dynamiques relationnelles qui sont à l'origine de la névrose du patient, et ce de manière à favoriser le développement de nouveaux modes de fonctionnement psychiques plus adaptés.

La cure psychanalytique classique s'opère selon un cadre strict et intensif : le client est allongé sur un divan et le thérapeute est assis derrière lui ; les séances de thérapie se déroulent à raison de trois fois par semaine en moyenne et la cure dure très longtemps,

parfois pendant des années. L'anonymat et l'opacité relatifs du thérapeute, ainsi que le fait qu'il est placé derrière le dos de la personne, servent à encourager l'installation de ce que Freud appelait une *névrose de transfert*, une actualisation des enjeux inconscients du patient dans le cadre de la relation thérapeutique, qui récapitulerait selon lui la névrose du patient. En analysant et en traitant cette névrose de transfert, la cure psychanalytique vise, par extension, à résoudre les conflits inconscients à l'origine de la névrose qui a motivé le patient à s'engager dans le traitement (Greenson, 1978). Ainsi, le focus principal de la cure psychanalytique se trouve dans l'interprétation et la résolution des résistances et du transfert, ce dernier étant conçu par Freud comme une forme de résistance, qui consiste à mettre inconsciemment en acte les motivations inconscientes plutôt que d'en prendre conscience et de les assimiler à sa personnalité.

1.2 La Neuropsychanalyse

Fondée il y a déjà vingt ans, la Neuropsychanalyse est une discipline intégrative qui emploie les méthodes de la neuropsychologie et des sciences cognitives afin vérifier, d'évaluer, et de modifier la conception de l'esprit offerte par la théorie psychanalytique. Si nous avons choisis de l'introduire à ce moment de notre exposé, c'est parce que la Neuropsychanalyse se concentre principalement sur l'investigation et la mise à jour des modèles théoriques de la psychanalyse freudienne. Ainsi, tout en les reformulant dans un langage théorique compatibles avec les neurosciences cognitives contemporaines, elle conserve certaines hypothèses de la psychanalyse freudienne qui ne seront pas ou peu représentés dans les modèles qui composent la psychologie psychodynamique. Ces hypothèses tournent autour du rôle central des affects, de la conscience (l'expérience subjective) et des émotions dans le fonctionnement psychologique humain, des

phénomènes qui auront une importance cruciale lorsque nous tenterons d'expliquer la psychothérapie psychodynamique à l'aide des neurosciences cognitives contemporaines.

Dans un article intitulé « The Conscious Id » (le Ça conscient), Mark Solms, le fondateur de la Neuropsychanalyse, a récemment procédé à une révision majeure de la psychanalyse freudienne à l'aide des connaissances actuelles dans le champ des neurosciences affectives et de la neuropsychologie de la conscience (Solms, 2013). Nous savons aujourd'hui que la conscience au sens le plus fondamental du terme, c'est-à-dire le fait de posséder une expérience subjective, dépend de structures sous-corticales situées dans la partie supérieure du tronc cérébral (Parvizi et Damasio, 2001). Cette partie du cerveau est associée au processus d'autorégulation biologique de l'organisme (homéostasie), ce qui a mené à la thèse selon laquelle la conscience serait issue de la modalité sensorielle de l'intéroception, qui traite des informations relatives à l'état du milieu interne du corps (Damasio, 2018). La conscience primaire décrite ici est conceptualisée comme étant composée d'états de conscience, qui sont caractérisés par différents niveaux (ex. éveil, sommeil), mais aussi par différentes qualités : des ressentis ou affects (Hobson *et al.*, 2000 ; Panksepp, 2005).

Dans ce modèle, les états affectifs résultent de l'intéroception, et forment la base de la conscience phénoménale en constituant le sujet de l'expérience. La valence des affects (agréable ou désagréable) représente la valeur (positive ou négative) des changements qui surviennent dans le milieu interne, vis-à-vis de la survie et du succès reproductif de l'organisme. Le milieu interne du corps fonctionne en grande partie de manière autonome, mais il active (excite) le système nerveux central à l'aide de signaux, les affects, qui guident l'allocation des ressources attentionnelles de manière à orienter la perception, l'action, et la cognition vers les stimuli pertinents pour le processus d'autorégulation de l'organisme. Ainsi, contrairement à ce que Freud pensait, la conscience dépendrait de structures sous-corticales primaires qui rendent compte des

besoins biologiques de l'organisme, une fonction qu'il associait plutôt à son concept d'Inconscient.

Là où le cortex entre en jeu, c'est dans la constitution des objets ou contenus de l'expérience consciente. Ceux-ci sont issus de la modalité sensorielle de l'extéroception, qui traite des informations sur le milieu externe de l'organisme. Cela reflète le fait que les besoins fondamentaux de l'organisme ne peuvent être satisfaits qu'à travers l'interaction avec le monde extérieur. Alors que les affects sont toujours conscients, puisqu'ils constituent la toile de fond de l'expérience consciente, l'extéroception peut tout à fait être inconsciente, tel que le montre la vaste littérature sur l'inconscient cognitif (Dehaene, 2014). Selon Solms, les objets constitués par l'extéroception doivent être investis par la conscience affective pour être expérimentés.

Cette affirmation rejoint directement la thèse selon laquelle la conscience est associée de manière consistante avec la mémoire de travail, une capacité permettant au système cognitif de maintenir actives des représentations extéroceptives (visuelles, auditives, etc.) afin de leur faire subir les opérations cognitives qu'exigent la tâche accomplie par l'agent. Ainsi, la mémoire de travail est au fondement des habiletés de cognition représentationnelle qui sous-tendent la capacité de pensée. En continuité avec la conception freudienne (cf. inhibition de l'action), la Neuropsychanalyse définit la pensée comme une forme de simulation virtuelle ou imaginaire des actions adaptatives que l'individu pourrait réaliser. Le but de ces expériences simulées serait d'évaluer, à l'aide de la valence homéostatique ressentie à travers l'affect, les conséquences potentielles de ces actions, et ce afin de pouvoir planifier et choisir les comportements les plus efficaces pour satisfaire les besoins biologiques de l'organisme. À partir de l'investissement de la conscience affective et de la mémoire de travail dans cette activité de cognition représentationnelle émerge une forme secondaire de conscience

(Edelman, 1992) la *conscience réflexive*³, à travers laquelle le sujet de l'expérience consciente se prend lui-même pour objet en se représentant ses propres états affectifs (et donc sa conscience primaire), ses comportements, etc. Bien que la distinction entre conscience primaire et secondaire n'était pas aussi claire dans ses écrits qu'elle ne l'est dans les sciences de l'esprits contemporaines, c'est à ce type de conscience secondaire que Freud référerait lorsqu'il affirmait que l'Ego était le siège de la conscience, et non pas à la conscience primaire (affective) mentionnée précédemment.

Solms insiste sur l'idée selon laquelle une partie importante de la cognition extéroceptive porte sur la représentation de l'état du corps de l'organisme, sur sa situation dans l'environnement, et vise à la régulation de son comportement ainsi que des états affectifs qui constituent l'expérience consciente du sujet. Solms identifie l'ensemble des circuits neurocognitifs qui sous-tendent cette activité d'autoreprésentation et d'autorégulation à l'Ego de la psychanalyse freudienne. Cependant, il affirme que ces processus cognitifs se déroulent en grande partie de manière inconsciente, et ne font qu'occasionnellement l'objet d'une prise de conscience réflexive. Comme nous l'avons vu précédemment, les affects sont ici conçus comme les signaux qui informent le système nerveux central des besoins de l'organisme, et qui motivent l'interaction avec l'environnement afin de les remplir adéquatement. Ce processus de motivation affective constitue la base de ces phénomènes fondamentaux que nous appelons les émotions, lesquelles sont conçues

³ La conscience secondaire est appelée différemment selon les disciplines et les orientations théoriques, et ce afin de capturer différents aspects des phénomènes complexes qui composent la catégorie générale de la conscience. On l'appelle par exemple la conscience déclarative, la conscience d'accès (*access consciousness*), la conscience autooëtique, la conscience étendue ou la pensée d'ordre supérieure. Nous avons choisi de privilégier le terme de conscience réflexive élaboré par Merleau-Ponty car il est communément utilisé au sein de la philosophie de l'esprit contemporaine.

ici comme des programmes neurocognitifs qui lient les affects avec des comportements qui ont pour but de répondre aux besoins qu'ils expriment.

Selon Solms, les phénomènes émotionnels constituent le cœur de la psychothérapie psychodynamique, et y sont considérés comme l'élément principal qui est en cause dans l'étiologie des psychopathologies qu'elle adresse. En effet, selon la théorie psychanalytique, le fonctionnement psychologique humain dérive du fonctionnement biologique de l'espèce humaine, et les besoins de base associés à ce dernier sont génétiquement déterminés, et donc innés, et fournissent les contraintes qui structurent le développement psychologique. Par conséquent, le développement psychologique est conçu comme le processus à travers lequel un organisme acquiert des structures comportementales et cognitives dynamiques en apprenant à remplir ses besoins fondamentaux dans son interaction avec son environnement. Si les émotions sont considérées comme l'élément central du fonctionnement psychologique humain, c'est parce qu'elles sont les événements mentaux qui reflètent directement le processus de satisfaction des besoins biologiques fondamentaux de l'organisme dans son interaction comportementale et cognitive avec son environnement. Ainsi, la souffrance émotionnelle qui caractérise les psychopathologies visées par la psychanalyse résulterait de l'incapacité d'un individu à remplir efficacement ses besoins dans son interaction avec son environnement.

En continuité avec la théorie psychanalytique, Solms affirme que comme la grande majorité des processus mentaux, la plupart de nos méthodes pour remplir nos besoins émotionnels sont exécutés inconsciemment. Cependant, et comme c'est aussi le cas pour l'ensemble de la psychologie psychodynamique contemporaine, la conception des processus mentaux inconscients en vigueur au sein de la Neuropsychanalyse est très proche de celle des sciences cognitives. En effet, à partir de la deuxième moitié du 20e siècle, la signification concept d'inconscient dynamique entretenue au sein de la psychologie psychodynamique s'est graduellement modifiée pour se rapprocher du

concept d'inconscient cognitif des sciences cognitives, et il est aujourd'hui très difficile de faire la distinction entre les deux notions (Burston, 1986; Kihlstrom, 1987; Stern, 1987)

Au sein des sciences cognitives contemporaines, l'inconscient cognitif est conçu comme un ensemble de processus de traitement de l'information ou de capacités comportementales qui se déroulent nécessairement de manière automatique ou qui ont été automatisés à travers un apprentissage (Kihlstrom, 1996). De plus, le caractère conscient ou inconscient des états mentaux est rarement abordé comme tel, mais l'on va plutôt parler de processus cognitifs explicites lorsque ceux-ci s'accompagnent d'une expérience consciente, et de processus cognitifs implicites pour ceux qui se déroulent automatiquement, et sans être accompagnés d'une expérience consciente. L'étude de la mémoire humaine a donné lieu à une autre distinction conceptuelle entre la mémoire déclarative qui implique la rétention d'unités informationnelles comme des faits ou des connaissances génériques, souvent par le biais du langage, et la mémoire procédurale qui implique des compétences ou des savoir-faire, qu'ils soient cognitifs ou comportementaux. La mémoire déclarative est considérée comme explicite, et la mémoire procédurale est considérée comme implicite. Cependant, ces deux types de mémoires sont en collaboration constante, et certaines connaissances procédurales peuvent être amenées à la conscience en les représentant sous une forme déclarative au moyen du langage ou d'un autre médium représentationnel.

À partir du cadre conceptuel des sciences cognitives, la Neuropsychanalyse redéfinit l'inconscient dynamique comme l'ensemble des processus cognitifs implicites et procéduraux qui sont impliqués dans le fonctionnement émotionnel d'un individu. Solms affirme que ce qui différencie la psychothérapie psychodynamique des autres formes de thérapie est qu'elle cherche à changer des modes de fonctionnement émotionnels fortement automatisés, inconscients, et qui sont donc consolidés sous une forme non déclarative qui ne permet pas qu'ils soient rappelés et transformés

directement dans la mémoire de travail (Solms, 2018). Comme tous les apprentissages procéduraux, ces modes de fonctionnement émotionnels inconscients seraient caractérisés par une importante rigidité, et l'un des objectifs principaux de la psychothérapie psychodynamique serait de mener l'individu à une prise de conscience réflexive de ces derniers afin de les modifier. Derrière cet objectif se cache l'hypothèse selon laquelle l'implication de la conscience réflexive rendrait possible une modification qualitative du fonctionnement émotionnel, qui se caractériserait par l'instauration d'un contrôle exécutif permettant une plus grande flexibilité des processus cognitifs qui le sous-tendent.

Selon Solms, le rôle de la pensée consciente serait de permettre la mise à jour des contenus mnésiques du système cognitif, un phénomène aujourd'hui appelé reconsolidation, et qui correspond au processus neuronal à travers lequel des souvenirs précédemment consolidés sont rendu malléables afin de pouvoir être modifiés (Arnold *et al.*, 2017). Adoptant une perspective fonctionnaliste, Solms conçoit les schémas cognitifs stockés dans la mémoire procédurale comme des apprentissages visant à résoudre, ou des solutions à, des problèmes concrets dans la vie de l'individu, en particulier en ce qui a trait à la satisfaction de ses besoins émotionnels fondamentaux. Cette perspective comprend la pensée comme une activité représentationnelle de simulation des situations concrètes auxquelles ces problèmes sont reliés afin d'élaborer de nouvelles solutions, et de permettre la modification des schémas mnésiques afin d'améliorer leur efficacité. C'est dans le même ordre d'idée qu'il propose de redéfinir le mécanisme du refoulement comme un processus d'automatisation prématurée de certains modes de fonctionnement émotionnels (Arnold *et al.*, 2017). Selon cette conception du refoulement, les schémas cognitifs inconscients ne sont pas activement maintenus en dehors de la conscience par une fonction défensive de l'Ego, mais ils ont simplement été consolidés sous une forme implicite et procédurale qui ne leur permet pas d'être rappelés et modifiés dans la mémoire de travail.

Solms explique cette automatisation précoce de modes de fonctionnement émotionnels inadaptés par le fait que l'attention et la mémoire de travail sont selon lui des ressources très limitées et que, lorsque présenté devant des problèmes insolubles, il est plus adaptatif pour le système cognitif d'automatiser une solution partielle ou inadéquate que de s'acharner à accomplir une tâche impossible. Ici, l'idée de problème insoluble vient remplacer la notion de conflit de la psychanalyse freudienne. En effet, un tel problème pourrait résulter de la contradiction insoluble entre certaines motivations instinctuelles (ex. agressivité) de l'individu et certaines exigences de son environnement (ex. normes sociales). Ce genre de situations d'impuissance face à des problèmes insolubles se produirait le plus souvent durant l'enfance, et cela associé au fait qu'il s'agit de la période développementale où se réalisent la majorité de nos apprentissages procéduraux durables, explique pourquoi une grande partie de nos modes de fonctionnement émotionnels inconscients trouveraient leur origine durant cette partie de notre vie.

En résumé, à partir d'une mise à jour de la psychanalyse freudienne, la Neuropsychanalyse nous offre un modèle explicatif de la psychothérapie psychodynamique qui est d'ores et déjà basé sur les connaissances actuelles en neurosciences affectives et cognitives contemporaines. Ainsi, selon la perspective Neuropsychanalytique, la tâche principale de la psychothérapie psychodynamique peut-être redéfinie comme la modification des schémas cognitifs procéduraux et implicites qui sous-tendent les modes de fonctionnements émotionnels inadaptés qui sont à l'origine de, ou ayant une influence sur, certaines formes de psychopathologie. En interprétant l'activité cognitive reflétée par les propos du patient (association libre), ainsi que son comportement dans la relation thérapeutique (transfert), le thérapeute a pour objectif principal de permettre au patient de relier son expérience affective (conscience primaire) aux automatismes cognitifs et comportementaux (schémas émotionnels inconscients) qu'ils occasionnent. Ce réinvestissement de la conscience primaire permettrait l'allocation des ressources attentionnelles et le recrutement de la

mémoire de travail afin de remplacer ces automatismes par une activité de pensée. Cette activité de pensée aurait pour but de permettre la reconsolidation de la mémoire procédurale afin de dé-automatiser les schémas émotionnels pathogènes, et d'élaborer de nouvelles solutions plus adaptatives aux problèmes dont ils constituaient une tentative de solution.

Du modèle explicatif de la psychothérapie psychodynamique résumé ci-dessus, nous pouvons extraire plusieurs hypothèses :

- La conscience et les affects, qui représentent la demande de travail endogène émergeant des besoins homéostatiques de l'organisme, exercent un pouvoir causal déterminant sur l'activité du système neurocognitif humain dans son ensemble, principalement à travers l'activation et la direction des ressources attentionnelles et des fonctions exécutives.
- Les émotions sont l'un des éléments centraux du fonctionnement psychologique d'un être humain, et ce car elles incarnent les processus neurocognitifs par le biais desquels ses besoins fondamentaux sont remplis à travers l'interaction comportementale et cognitive avec le monde extérieur. Les émotions seraient acquises à travers l'expérience durant le développement et au-delà, et pourraient donc être modifiées dans le cadre de la psychothérapie psychodynamique.
- La grande majorité des processus cognitifs qui sous-tendent le fonctionnement psychologique humain, ce qui inclue les émotions, se déroulent de manière inconsciente, ce qui veut dire : a) qu'ils sont implicites, c'est-à-dire qu'ils ne s'accompagnent pas d'une expérience subjective dont ils sont l'objet ; et b) qu'ils sont automatiques, c'est-à-dire qu'ils n'ont pas besoin de faire l'objet d'un investissement des ressources attentionnelles ou des fonctions exécutives pour être implémentés.

Autrement dit, ils se déroulent en dehors de la conscience et du contrôle volontaire de l'individu.

- Un sous-ensemble non-négligeable des problèmes de santé mentale qui affectent les êtres humains auraient pour facteur étiologique, principal ou secondaire, la prédominance de processus émotionnels inconscients inadaptés qui, en empêchant l'individu de remplir efficacement ses besoins dans son interaction avec son environnement, seraient à l'origine d'une détresse psychologique importante. Ce sont ces pathologies mentales qui sont visées par la psychothérapie psychodynamique.

- La direction de la conscience affective, par le biais de l'investissement des ressources attentionnelles, sur les manifestations cognitives et comportementales qui émergent de processus émotionnels automatiques inadaptés, permettrait la mise en place d'un contrôle exécutif autorisant leur modification à travers la reconsolidation des schémas neurocognitifs qui les sous-tendent. C'est à travers ce processus de prise de conscience (réflexive) que la psychothérapie psychodynamique cherche à modifier les modes de fonctionnement émotionnels inadaptés qui seraient à l'origine des formes de psychopathologies qu'elle adresse.

1.3 La psychologie psychodynamique

La psychothérapie psychodynamique contemporaine repose sur un cadre théorique élargit, auquel nous référerons dans notre mémoire à l'aide de l'appellation générale de « psychologie psychodynamique », et qui comprend un ensemble de modèles théoriques qui, bien que trouvant leur origine dans la psychanalyse freudienne, ont fait

grandement évoluer la théorie psychanalytique, en particulier durant la seconde moitié du 20^e siècle. L'ensemble des modèles théoriques qui forment la base à partir de laquelle s'est développé la psychologie psychodynamique comprend minimalement (Gabbard et Crocq, 2010) : la psychanalyse freudienne, la théorie des relations d'objet (Fairbairn, 1986 ; Klein, 1999 ; Winnicott, 2005), la théorie de l'attachement (Ainsworth *et al.*, 1978 ; Bowlby, 1999), la psychologie du moi (Freud, 1992 ; Hartmann, 1964 ; Rapaport, 1967) et la psychologie du Self (Kohut, 2009). Le développement de la psychologie psychodynamique a émergé du besoin ressenti par une communauté grandissante de psychothérapeutes d'affiner, d'améliorer et d'étendre le modèle du fonctionnement mental humain proposé par la théorie psychanalytique afin de mieux rendre compte des phénomènes en jeu au sein de la pratique de la psychothérapie. L'élaboration du cadre théorique intégratif de la psychologie psychodynamique est caractérisé par la théorisation de deux axes thématiques principaux qui sont aujourd'hui considérés comme les deux domaines de phénomènes fondamentaux adressés par la psychothérapie psychodynamique : 1) le rôle déterminant des relations interpersonnelles dans l'acquisition et la régulation du fonctionnement psychologique humain ; et 2) l'importance critique des fonctions associées à l'Ego, aujourd'hui généralement appelé le Soi (« self » en anglais), ainsi que du processus développemental qui mène à leur acquisition.

La théorie des pulsions affirme que les motivations instinctuelles intrinsèques à la structure biologique de l'espèce humaine vont se manifester dans la recherche d'objets externes concrets permettant d'assouvir les besoins qu'elles expriment. La théorie des relations d'objets, initialement développée par Mélanie Klein, associe cette prémisse au constat suivant : durant les premières années de vie d'un individu, ses pulsions sont généralement satisfaites systématiquement par les mêmes objets, d'autres individus, ses donneurs de soins primaires (souvent ses parents). De plus, tout au long de la vie d'un être humain, les autres individus qui l'entourent gardent un rôle critique dans la satisfaction de ses besoins fondamentaux. La théorie des relations d'objet part de ce

constat général pour proposer un modèle du fonctionnement mental humain comme étant principalement concerné par, et organisé autour des interactions sociales essentielles à la satisfaction de ces pulsions. Ce faisant, la théorie des relations d'objet consiste en un modèle développemental qui affirme que les expériences précoces dans le contexte de l'interaction entre soi et autrui, et les états affectifs qui leur sont associés sont intériorisés pour créer des représentations internes des relations interpersonnelles. Ces représentations formeraient la base de croyances qui portent à la fois sur les autres, les « objets » de l'expérience, et sur soi-même, le soi, le « sujet » de l'expérience, ainsi que sur leur interaction. Ces croyances seraient ensuite appliquées à toutes les relations interpersonnelles au cours de la vie de l'individu, et les problèmes émotionnels vécus par le patient et partagés dans la psychothérapie sont conçus comme résultant des difficultés relationnelles vécues durant l'enfance.

La ligne de pensée ouverte par la théorie des relations d'objet a trouvé son expression la plus marquante dans les travaux du psychiatre anglais John Bowlby (1969) et sa fameuse théorie de l'attachement, qui théorise le caractère inné des besoins sociaux de l'être humain (Bowlby, 1999). La théorie de l'attachement part du constat selon lequel l'espèce humaine est une espèce *altricielle*, ce qui veut dire que le nouveau-né humain est typiquement prématuré, incompetent, et vulnérable, et qu'il est dépendant de l'aide de son environnement social pour assouvir ses besoins biologiques fondamentaux, et ce pendant une période développementale particulièrement longue (Sleigh, 2011). Selon la théorie de l'attachement, chaque enfant est génétiquement prédisposé à chercher une relation de dépendance et de support émotionnel avec un donneur de soin primaire (généralement un parent), envers lequel il se tourne pour trouver confort et sécurité face au stress et aux menaces perçues. La qualité de ces relations, en particulier dans la jeune enfance, mène au développement de schémas d'attachements particuliers, qui résultent à l'âge adulte en des *styles d'attachement* regroupés en quatre catégories : sécuritaire, anxieux-soucieux, distants-évitant, craintif-évitant (ambivalent ; pour plus d'information, voir Cassidy & Shaver, 2002; Fraley & Roisman, 2019). Ceux-ci

correspondent à un modèle interne de représentation du niveau de confiance accordé au support potentiel du donneur de soins dans la tâche de régulation homéostatique et émotionnelle de l'enfant.

Lors d'épisodes de stress ou de dérégulation homéostatique, les différents types d'attachement se manifestent par des comportements différents. Les interactions précoces sont conçues comme le déterminant fondamental du développement des capacités d'autorégulation des états physiologiques et affectifs, des habiletés qui sont ancrées dans des structures cérébrales déterminées par une plasticité dépendante de l'expérience et l'épigénétique. Le type d'attachement qui en résulte est généralisé vers les autres relations et situations sociales rencontrées plus tard dans la vie de l'individu (Cittern *et al.*, 2018).

Les deux approches introduites ci-dessus forment le point de départ de ce que (Mitchell, 2000) a nommé le tournant relationnel de la psychanalyse, et qui correspond à un changement de perspective consistant à mettre l'emphase sur le rôle fondamental des relations interpersonnelles dans la constitution de l'esprit humain. Ce tournant est basé sur l'adoption de deux thèses connexes, mais distinctes, sur la nature sociale de l'être humain. La première, évidente dans les travaux de Bowlby, consiste à affirmer que nos besoins sociaux sont intrinsèques et innés et qu'ils n'émergent pas durant le développement, tel que le pensait Freud, comme un simple moyen d'assouvir nos besoins biologiques fondamentaux. La seconde consiste à affirmer que notre esprit est socialement construit, c'est-à-dire que le fonctionnement mental typiquement humain serait le fruit d'un conditionnement social qui s'opère durant le développement à travers les relations interpersonnelles.

L'adoption de ces axiomes va amener une modification significative du modèle de l'esprit proposé par la psychanalyse freudienne. En effet, dans la psychanalyse freudienne, les unités de base de l'esprit sont les contenus mentaux issus des

motivations instinctuelles que sont les pulsions (désirs, souhaits, intentions, etc.). Au sein de la psychologie psychodynamique contemporaine, les unités de base de l'esprit sont des formes de modèles internes des dynamiques relationnelles qui déterminent les interactions entre un individu et les autres êtres humains qui composent son environnement social. Cette conception de l'esprit émerge de l'adoption d'une perspective développementale qui affirme que les interactions répétées qui ont lieu, en particulier durant les premières années de vie, entre un individu et ses donneurs de soins primaires, mènent à la formation de représentations qui constituent des abstractions ou des généralisations d'interactions prototypiques. Ces représentations ont reçu différentes appellations selon les auteurs : représentations d'interactions généralisées (RIGs ; Stern, 1998) ; structures interactionnelles (Beebe *et al.*, 1997) ; schémas (*patterns*) d'attachement (Ainsworth *et al.*, 1978) ; configurations relationnelles (Mitchell, 1988), etc.

Au sein de la psychologie psychodynamique, la partie inconsciente du fonctionnement mental est conçue comme étant essentiellement composée de représentations interactionnelles de ce type. L'influence des sciences cognitives a mené à la compréhension des processus mentaux inconscients comme une forme de connaissance procédurale portant sur les interactions sociales, un *savoir-faire relationnel implicite* acquis durant le développement et qui est appliqué systématiquement à l'ensemble des relations interpersonnelles entretenues par l'individu tout au long de sa vie (Lyons-Ruth *et al.*, 1998). Ce sont ces mêmes schémas relationnels inconscients qui sont conçus comme étant responsables de l'émergence des formes de psychopathologies adressées par la psychothérapie psychodynamique. En effet, le constructivisme social qui caractérise la conception de l'esprit développée par la psychologie psychodynamique débouche en un modèle de la psychopathologie comme résultant d'un *échec environnemental* où les relations avec les donneurs de soins primaires durant le développement sont considérées comme le facteur étiologique principal des troubles mentaux adressés par la psychothérapie psychodynamique. Dans ce contexte, les

psychopathologies émergeraient de l'apprentissage et de l'application systématique de modes de fonctionnement relationnels qui ne sont pas ou ne sont plus adaptatifs.

Cependant, le rôle des relations interpersonnelles dans le développement d'un fonctionnement mental sain ou pathologique ne s'arrête pas là, ce qui nous amène au deuxième axe thématique de la psychologie psychodynamique. En effet, une partie importante des modèles théoriques qui composent la psychologie psychodynamique stipulent que notre dépendance prolongée à notre environnement social joue un rôle critique dans l'acquisition de fonctions psychologiques essentielles à la santé mentale d'individu, et qui correspondent aux fonctions que Freud attribuait à sa structure de l'Ego.

Cette perspective émerge de la psychologie de l'Ego, un courant théorique initié par les travaux d'Anna Freud, de Heinz Hartmann, et de David Rapaport, et qui considérait l'ensemble des fonctions de l'Ego comme les déterminant principaux du fonctionnement psychologique et de la santé mentale d'un individu. Au-delà des fonctions défensives sur lesquelles Freud se concentrait, les partisans de la psychologie de l'Ego ont mis l'accent sur le rôle critique d'habiletés telles que : le contrôle de l'impulsivité permettant d'inhiber les comportements excessifs, dangereux, et socialement inacceptables (ex. violence, promiscuité sexuelle, etc.) ; la capacité de distinguer ce qui appartient au monde extérieur de ce qui est de l'ordre des phénomènes mentaux ; la raison et le jugement, permettant de penser et d'agir de manière organisée, efficace, et adaptative ; la capacité de réguler son expérience affective, etc.

Dans la même ligne de pensée, un autre modèle théorique influent est celui de la psychologie du *Self* développé par Ernst Kohut, dans laquelle le fonctionnement et la santé mentale sont compris comme étant principalement déterminés par le développement d'une représentation de soi-même qui soit à la fois positive et réaliste. Le *Self*, aujourd'hui communément appelé *concept de soi*, est une représentation

synthétique qu'un individu entretient à propos de sa personne dans son ensemble, et qui servirait à évaluer et à réguler ses états mentaux, affectifs et comportementaux. Kohut affirme que le bien-être psychologique d'un individu dépend de sa capacité à maintenir un amour propre (appelé narcissisme en psychanalyse) suffisamment élevé tout en gardant une vision réaliste de sa personne qui lui permette d'être fonctionnel et d'entretenir des relations saines. Ici, le *Self* est conçu comme résultant entièrement la qualité des relations interpersonnelles durant le développement de l'individu. L'apport central de la psychologie du *Self* à la psychologie psychodynamique réside dans la définition des caractéristiques communes que doivent démontrer un donneur de soin ou un thérapeute lors de leur interaction avec un individu afin d'éliciter le développement d'un concept de soi sain.

Dans ce contexte, les donneurs de soins primaires, ou toute autre personne jouant un rôle indispensable dans la vie d'un individu, sont théorisés comme faisant partie intégrante du *Self* de cet individu, une indifférenciation relative qui est conçue à travers la métaphore biologique de la symbiose (Mahler *et al.*, 1975). Kohut réfère à ces personnes à travers l'appellation de *selfobjects*, et c'est leur capacité d'accomplir adéquatement certaines fonctions critiques pour l'individu qui détermine selon lui le développement d'un *Self* sain chez ce dernier (Kohut, 2009). La première de ces fonctions est l'empathie, la capacité de comprendre et/ou de ressentir ce que l'individu expérimente en se mettant à sa place. La deuxième de ces fonctions, indissociable de la première, est le *mirroring*, qui consiste à mimer le comportement affectif de l'individu afin de le lui refléter. Réalisées conjointement, ces fonctions vont tout d'abord permettre à l'individu se sentir compris, soutenu, et accompagné dans son expérience, ce qui devrait le pousser à développer une confiance en son donneur de soin, et par extension, en lui-même. De plus, Kohut considère l'empathie comme une forme d'introspection vicariante permettant au donneur de soin de participer directement au développement de la conception de soi que l'individu se fait de lui-même (Agosta, 2010).

Ensemble, la psychologie de l'Ego et la psychologie du Self ont contribué à placer le phénomène du Soi psychologique au centre de la conception de l'esprit de la psychologie psychodynamique. Selon cette perspective, les interactions sociales durant le développement de l'individu n'ont pas seulement pour effet de déterminer ses habiletés de cognition sociale et ses automatismes relationnels, mais permettent aussi la transmission de certaines fonctions psychologiques fondamentales de l'organisation de l'esprit humain. Au sein de la psychologie psychodynamique contemporaine, ces habiletés d'autorégulation affectives, cognitives, et comportementales sont conçues comme étant intrinsèquement reliées au développement d'un concept de soi positif et réaliste à travers l'interaction avec des donneurs de soins attentifs et empathiques.

1.4 La théorie de la mentalisation

Issue de l'intégration de la théorie de l'attachement, de la théorie psychanalytique, et de la psychologie développementale, la théorie de la mentalisation (TM), développée par Peter Fonagy et ses collègues (Fonagy *et al.*, 2018) est le modèle théorique contemporain le plus abouti en ce qui concerne la compréhension du concept psychodynamique de Soi (ou Ego), des fonctions cognitives qui lui sont attribuées, et de la manière dont ces dernières sont acquises durant l'ontogénie à travers les relations interpersonnelles. Les partisans de la théorie de la mentalisation adoptent la perspective de la théorie de l'esprit afin d'expliquer le développement émotionnel. Ce faisant, ils définissent la capacité de mentalisation (ou théorie de l'esprit) comme une activité mentale imaginative à propos des autres ou de soi-même, et qui implique de percevoir et d'interpréter le comportement et les états affectifs humains en termes d'états mentaux intentionnels (besoins, désirs, sentiments, croyances, objectifs, buts et

raisons). Selon eux, il est clair que les émotions sont aussi des états mentaux intentionnels que nous attribuons aux « esprits » des individus qui composent notre environnement social (ainsi qu'à nous-mêmes) pour expliquer et prédire leurs comportements.

Selon la théorie de la mentalisation, une émotion est un état mental spécifique en ce qu'il est composé de trois dimensions fondamentales (Fonagy *et al.*, 2018). Premièrement, une émotion contient une certaine disposition à l'agent visé, appelé contenu (ou information) dispositionnel(le), qui réfère aux comportements probables, à la tendance à agir d'une certaine façon, d'un agent se trouvant dans un état émotif particulier. Deuxièmement, une émotion implique une expérience subjective caractéristique, un état affectif qui reflète des changements physiologiques dans le corps de l'individu, ce qui facilite la précision lors de l'autoattribution. Troisièmement, les émotions sont accompagnées par des expressions faciales saillantes et spécifiques qui pourraient les rendre plus faciles à reconnaître chez autrui que chez soi-même. Selon les partisans de la TM, l'ensemble des indices intéroceptifs et proprioceptifs qui sont activés lors d'un événement émotif et de son expression ne sont, dans un premier temps, pas perçus consciemment par l'enfant, et ne sont pas regroupés catégoriquement de telle manière qu'ils puissent être appréhendés comme un état émotif distinct. La question devient alors de savoir comment l'enfant en vient à prendre conscience et à se représenter ces ensembles d'indices sur son état interne comme des marqueurs d'état émotifs catégoriquement distincts.

Pour répondre à cette question, la TM adopte une conception dite « biosociale » du développement émotif (LaFreniere, 2000), qui affirme que dès la naissance, la mère et l'enfant forment un système de communication affective dans lequel la mère joue un rôle interactif vital dans la modulation des états affectifs (et homéostatiques) de l'enfant. La TM propose que la propension, spécifique à l'espèce humaine, au reflet facial et vocal de l'expressivité émotive de l'enfant durant ces interactions de

régulation affective joue un rôle crucial dans le processus développemental permettant l'acquisition de la capacité de mentalisation, conçue comme le mécanisme central qui sous-tend l'autorégulation émotionnelle.

Au départ, le comportement de l'enfant serait gouverné par des « émotions primaires » (Panksepp, 2006, 2010) conçues comme des automatismes affectifs comportementaux sur lesquels il n'a pas de contrôle dans un premier temps. La régulation affective est effectuée principalement par le donneur de soins qui, lisant les expressions émotionnelles automatiques de l'enfant, y réagit avec les interactions de modulation affective appropriées. Les interventions performées par le parent peuvent être divisées en deux dimensions principales, qui vont le plus souvent être appliquées en parallèle.

Le premier type d'intervention consiste à répondre au besoin homéostatique correspondant à l'état affectif reconnu chez l'enfant : le nourrir s'il a faim, le réchauffer s'il a froid, etc. Le deuxième type d'intervention, celui qui nous intéresse le plus, correspond à une activité de communication au sein de laquelle il y a transmission du parent vers l'enfant de l'interprétation faite (par le système cognitif du parent) de l'ensemble des indices affectifs, faciaux, et comportementaux qu'il démontre comme faisant partie d'un état mental unitaire : une émotion. Pour ce faire, il faut tout d'abord aider l'enfant à prendre conscience de l'ensemble de ces indices. Cela passe par exemple par la manière dont le parent va mimer les expressions faciales de l'enfant, de manière à lui fournir un reflet extéroceptif de cet aspect de son état auquel il n'aurait pas accès autrement, et qu'il va pouvoir associer aux sensations proprioceptives correspondantes (Gergely, 2007). Un autre aspect important est l'introduction par le parent de représentations auditives, des mots, systématiquement associés à certains états émotionnels qu'il reconnaît chez l'enfant. À travers la répétition de ces interactions de modulation affective, l'enfant va commencer à intérioriser l'interprétation que son parent fait de son expérience et va développer une compréhension des concepts émotionnels employés par ce dernier. Les mots vont servir

d'outils représentationnels permettant de ramener ces concepts à la mémoire de travail dès que l'ensemble des indices qui lui correspondent se manifestent. L'idée est que cette capacité d'interpréter son expérience et ses comportements à l'aide de concepts mentaux et émotionnels est le mécanisme principal de l'autorégulation émotionnelle.

Selon la TM, l'autocontrôle émotionnel va devenir possible avec l'établissement de structures de contrôle secondaires qui a) suivent, détectent, et évaluent les changements d'états affectifs de l'organisme ainsi que les manifestations comportementales qu'ils engendrent. ; et b) peuvent inhiber ou modifier la réaction émotionnelle automatique si ses conséquences anticipées rentrent en conflit avec les plans cognitifs de niveau supérieur. L'important dans ce processus est la possibilité pour l'individu en développement d'apprendre à se représenter le contenu dispositionnel de ses états émotionnels sous une forme déclarative, et ce de manière à ce qu'il soit accessible à la conscience réflexive, ce qui va permettre l'évaluation et la régulation des comportements qui vont émerger de ses états affectifs. Cette même capacité va être appliquée simultanément à la compréhension des comportements des autres individus en leur attribuant, à eux aussi, des états mentaux intentionnels, et va devenir l'un des piliers de la cognition sociale, permettant à l'individu d'anticiper le déroulement des interactions sociales et de développer des relations interpersonnelles saines et durables.

En plus de sous-tendre l'acquisition des capacités d'autorégulation émotionnelles et de cognition sociale mentionnées précédemment, le développement de la capacité de mentalisation devrait permettre le développement du *concept de soi*, un complexe de schéma cognitifs qui portent sur la description et l'évaluation de l'individu lui-même, et qui sous-tendraient l'acquisition et le maintien d'un sens de *l'identité* à travers le temps (Hattie, 2014). Autrement dit, selon la TM, le processus développemental qui permet l'acquisition du fonctionnement émotionnel est le même que celui qui mène à l'émergence de ce que l'on appelle le *Soi psychologique*. Ce dernier est ici conçu comme un complexe développemental unitaire composé de l'ensemble des schémas

cognitifs qui composent le concept de soi, ainsi que de la capacité de mentalisation qui sous-tend leur apprentissage et leur utilisation dans le cadre de l'autorégulation émotionnelle et de la cognition sociale. Si nous parlons ici du développement du *Soi psychologique*, c'est parce que nous avons affaire au processus à travers lequel un individu en vient à se considérer lui-même comme un agent intentionnel sujet à des états mentaux, et à comprendre à réguler ses états et ses actions selon ce modèle qu'il a de lui-même.

Avec la progression du développement cognitif, l'individu va se mettre à adopter des concepts de plus en plus englobants, pour se représenter des dimensions de plus en plus abstraites et complexes de son état sur des étendues temporelles de plus en plus larges. C'est ainsi que l'individu va se mettre à utiliser des concepts comme l'identité ou la personnalité pour se représenter les régularités de son fonctionnement mental et les patterns d'interaction récurrents qu'il entretient avec son environnement social. De plus, l'individu va s'ouvrir progressivement à l'influence d'un environnement social étendu qui va comprendre l'ensemble des interactions interpersonnelles auxquelles il participe où qu'il observe. Néanmoins, les relations interpersonnelles intimes entretenues avec ses figures d'attachement principales vont continuer d'avoir une influence déterminante dans le fonctionnement mental et émotionnel de l'individu, ainsi que dans les croyances qu'il entretient à propos de lui-même.

À partir d'une reformulation des théories de la psychologie psychodynamique dans le langage conceptuel de la psychologie cognitive et développementale contemporaine, la théorie de la mentalisation nous permet de compléter notre cadre théorique initial à l'aide des hypothèses suivantes :

- Le fonctionnement psychologique typiquement humain, ce qui inclue les émotions, est acquit par le biais d'un processus développemental qui dépend en grande partie d'un conditionnement social qui s'opère durant l'ontogénie et au-delà, à travers

les relations interpersonnelles qu'un individu entretient avec les autres êtres humains qui composent son environnement social proximal.

- Le développement des capacités d'autorégulation émotionnelles repose principalement sur l'apprentissage, à travers les interactions sociales, de la théorie de l'esprit (ou mentalisation), une habileté métacognitive lui permettant de percevoir et de se représenter ses états affectifs et les activités cognitives ou comportementales qui en résultent comme étant causalement explicables de par leur appartenance à des états de faits unitaires dont il est le sujet, des états mentaux intentionnels particuliers que l'on appelle des émotions. Cette habileté sous-tendrait aussi une grande partie des capacités de cognition sociale essentielles au mode de vie humain, permettant à un individu d'interagir efficacement avec, et de s'intégrer à, son environnement social, ainsi que d'entretenir des relations interpersonnelles saines et durables.

- L'acquisition de la capacité de mentalisation marquerait aussi l'avènement d'une nouvelle forme d'agentivité au sein de laquelle un être humain est capable de se représenter lui-même en tant qu'individu, et ce grâce au développement d'un concept de soi, à travers lequel il se conçoit comme une entité à part entière (différenciée d'autrui) qui présente ses propres caractéristiques, et comme un agent intentionnel sujet à des états mentaux (comme un esprit). Tout comme les différents concepts émotionnels, le concept de soi jouerait un rôle important dans l'autorégulation émotionnelle car il serait utilisé par l'individu pour expliquer et réguler ses états affectifs, cognitifs, et comportementaux. Le complexe développemental et fonctionnel formé par la capacité de mentalisation et les schémas émotionnels qui en émergent (incluant le concept de soi), est ici conçu comme correspondant directement la notion d'Ego de la psychanalyse Freudienne, capturée aujourd'hui à travers le concept de Soi psychologique. Ce sont ces fonctions et ces schémas cognitifs qui seraient l'objet de la psychothérapie psychodynamique.

- Le développement de la capacité de mentalisation étant dépendante de la qualité des interactions sociales avec les donneurs de soins primaires (figures d'attachement) durant l'ontogénie, l'échec de ces dernières à soutenir son apprentissage serait le facteur étiologique principal des modes de fonctionnement émotionnels inconscients inadaptés qui caractérisent les pathologies mentales adressées par la psychothérapie psychodynamique. Ces dernières seraient dues à un dysfonctionnement et/ou à une déficience de la capacité de mentalisation qui mèneraient respectivement : a) à la formation et à l'implémentation continuelle de schéma émotionnels et d'une représentation de soi erronés et pathogènes ; et/ou b) à une incapacité relative à réguler ses états affectifs et comportementaux en se les représentant comme des états mentaux intentionnels, et à maintenir un concept de soi cohérent qui puisse permettre de maintenir un sens de l'identité à travers le temps.

- La psychothérapie psychodynamique correspondrait à une tentative de recréer les conditions d'acquisitions idéales de la capacité de mentalisation, une relation interpersonnelle saine avec une figure d'attachement sécuritaire, et ce afin de favoriser son apprentissage adéquat, ainsi que le développement de schémas émotionnels, d'habileté de cognition sociale et d'un concept de soi fonctionnels favorisant le bien être psychologique de l'individu et son intégration à son environnement social.

1.5 La psychothérapie psychodynamique

En plus d'être basée sur le cadre théorique élargi présenté précédemment, la psychothérapie psychodynamique contemporaine se différencie de la cure psychanalytique à travers la modification du cadre clinique qui caractérise sa méthode

d'intervention. En effet, afin de maximiser l'accessibilité de la psychothérapie à l'ensemble des profils psychologiques, et de réduire les coûts du traitement, la psychothérapie psychodynamique adopte une approche plus flexible permettant d'adapter le cadre clinique, l'intensité et la durée du traitement aux besoins du client. La psychothérapie psychodynamique se déroule en face-à-face, à raison d'une fois par semaine en moyenne, et peut-être aussi brève qu'une quinzaine de sessions dépendamment des besoins du patient. Néanmoins, les traitements de longue durée (plus de six mois) restent recommandés, et sont l'une des caractéristiques distinctives de la psychothérapie psychodynamique par rapport à d'autres formes de psychothérapie.

Le terme de psychothérapie psychodynamique est aujourd'hui utilisé pour référer à un ensemble de pratiques psychothérapeutiques situées sur un continuum entre deux pôles qui caractérisent l'objectif du traitement : l'interprétation et/ou le soutien (Leichsenring *et al.*, 2006). Ces deux pôles se concentrent sur deux aspects différents de la conception du processus de changement psychologique dans la psychothérapie psychodynamique.

Une thérapie axée sur l'interprétation aura pour but d'améliorer la compréhension que le patient a de son fonctionnement émotionnel et relationnel, et ce afin de faciliter sa modification. Ce type de thérapie est fondé sur l'engagement du patient et du thérapeute dans une pratique collaborative d'interprétation de l'activité cognitive spontanée du patient (cf. association libre), et qui repose sur l'hypothèse d'un déterminisme psychologique qui stipule que la pensée du patient va : a) se diriger naturellement vers les éléments qui présentent la plus grande valence émotionnelle pour lui ; et b) démontrer la manière dont il comprend et se représente son expérience affective, ses états mentaux, et ses relations interpersonnelles. Le patient va parler le plus souvent des expériences émotionnelles qui émergent de sa vie relationnelle, à la fois présente et passée. Si cela est possible, l'un des objectifs principaux sera d'identifier, dans le passé relationnel de l'individu, les expériences marquantes qui ont joué un rôle

déterminant dans la mise en place du fonctionnement émotionnel et relationnel actuel du patient. Cependant, comme ce fonctionnement est majoritairement de nature inconsciente et procédurale, il est fort commun que le patient ne puisse pas s'en rappeler directement.

C'est ainsi qu'entre en scène l'activité qui est au cœur de la psychothérapie psychodynamique : l'interprétation du transfert. Le transfert, qui peut être défini comme la répétition des expériences passées dans les relations interpersonnelles actuelles, consiste en des régularités émotionnelles et comportementales qui émergent d'expériences ontogénétiques précoces. Selon la perspective Neuropsychanalytique, ces modes de fonctionnement émotionnels sont considérés comme des apprentissages procéduraux réalisés durant le passé développemental de l'individu, qui ont été encodés sous la forme de schémas cognitifs implicites et automatiques, dont l'implémentation s'effectue en dehors de la conscience et en l'absence du contrôle volontaire de l'individu. L'interprétation du transfert correspond au processus à travers lequel le psychothérapeute guide le patient vers une prise de conscience progressive de ses modes de fonctionnement émotionnels et relationnels actuels, en l'aidant à reconnaître leurs manifestations comportementales et psychologiques récurrentes dans la relation thérapeutique. Cette prise de conscience, à travers la direction des ressources attentionnelles du patient, permettrait la mise en place d'un contrôle exécutif permettant la modification des schémas émotionnels par le biais du mécanisme de reconsolidation neuronale. En résumé, à travers l'application des techniques de l'association libre et de l'interprétation du transfert, une thérapie axée sur l'interprétation vise à identifier directement les dynamiques affectives, cognitives et comportementales qui sous-tendent le fonctionnement émotionnel et relationnel effectif du patient.

Une thérapie axée sur le soutien aura pour but d'assister et d'entraîner le patient à l'exercice d'habiletés cognitives d'autorégulation émotionnelle et de cognition sociale

à qui sont essentielles au maintien de sa santé mentale, soit parce qu'elles sont rendues temporairement inaccessibles à cause d'un stress aigu (ex. événement traumatique), ou qu'elles n'ont pas été suffisamment développées durant l'ontogénie. Il a été proposé d'identifier le processus de changement correspondant à une thérapie axée sur le soutien au développement ou au maintien des fonctions de l'Ego ou du Soi (Wallerstein, 1989), un complexe psychologique qui serait notamment responsable de la constitution du concept de soi à travers lequel un individu construit son identité à travers le temps. Selon la théorie de la mentalisation, l'ensemble de ces fonctions reposent sur la capacité de mentalisation, une faculté métacognitive permettant à l'individu de percevoir et de se représenter ses états affectifs et les activités cognitives ou comportementales qui en résultent comme étant causalement explicables de par leur appartenance à des états de fait unitaires dont il est le sujet, des états mentaux intentionnels particuliers que l'on appelle des émotions. D'après cette perspective, cette faculté résulterait d'un apprentissage réalisé dans le cadre des relations d'attachement qui unissent un individu et ses donneurs de soins primaires durant le développement.

Dans ce contexte, l'objectif d'une thérapie axée sur le soutien consiste à tenter de reproduire le cadre d'une relation d'attachement sécuritaire au sein de laquelle le psychothérapeute va pouvoir entraîner le patient à l'exercice de sa capacité de mentalisation. Cette transmission passe par un travail continu et répété de collaboration dans lequel le thérapeute aide le patient à mentaliser ses états émotionnels, démontrant ainsi les fonctions cognitives adéquates afin que ce dernier puisse se les approprier progressivement. Dans un premier temps, il est possible que le thérapeute reprenne le rôle du donneur de soin primaire en réalisant, au moins en partie, les fonctions de régulation affective et de mentalisation à la place du patient. Par la suite, les expériences vécues dans la relation effective avec le thérapeute vont fournir une nouvelle base de souvenirs épisodiques (mémoire déclarative) sur lesquels ce dernier va pouvoir se reposer pour se représenter son fonctionnement mental, émotionnel et relationnel. C'est à partir de l'amélioration et de la solidification de ces fonctions que

le patient va pouvoir progressivement les appliquer afin de modifier effectivement son fonctionnement émotionnel.

Dans les faits, la grande majorité des psychothérapies psychodynamiques utilisent simultanément des techniques d'intervention et de soutien afin de viser à la fois l'identification du fonctionnement émotionnel et relationnel actuel et l'amélioration des fonctions d'autorégulation du Soi (psychologique) qui permettront de le modifier. Ce processus de changement visé par la psychothérapie psychodynamique est souvent accueilli de manière ambivalente par le patient. L'équilibre psychologique du patient a été mis en place à travers des années d'exercice de schéma émotionnels utilisés pour éviter l'expérience d'affects douloureux. La psychothérapie psychodynamique vise précisément à déconstruire cet équilibre dysfonctionnel, et le patient peut inconsciemment s'opposer aux efforts du thérapeute visant à la prise de conscience et au changement de son fonctionnement émotionnel. Ces tendances à l'inertie, ou à la préservation du statut quo, qui émergent en réaction aux expériences émotionnelles désagréables élicitées par la thérapie sont appelés des résistances, et leur interprétation, afin de les surmonter, est l'un des mécanismes principaux du changement dans la psychothérapie psychodynamique. Cela peut être compris comme un processus de désautomatisation des schéma émotionnels initiaux, qui passe notamment par la constitution, à travers l'exercice de la capacité de mentalisation, de nouveaux schémas plus efficaces.

Pour finir, nous pouvons définir la psychothérapie psychodynamique comme une modalité d'intervention psychosociale qui consiste à mettre en place une relation interpersonnelle aux qualités spécifiques (ex. attachement sécuritaire) entre le patient et un psychothérapeute entraîné, et ce afin de recréer les conditions d'acquisition du fonctionnement émotionnel et relationnel du patient dans le but de le modifier. Cette modification passe par deux mécanismes principaux : a) la prise de conscience, par le patient, de son fonctionnement émotionnel et relationnel, principalement à travers

l'interprétation de ses manifestations actuelles dans son activité cognitive (cf. association libre) et dans la relation avec le psychothérapeute (transfert); et b) l'acquisition, à travers les expériences vécues dans la relation effective avec le thérapeute, de capacités d'autorégulation émotionnelles (ex. mentalisation) auparavant déficientes ou insuffisantes, qui vont lui permettre de modifier son fonctionnement émotionnel.

CHAPITRE II

EXPLANANS : UN MODÈLE DE L'ESPRIT HUMAIN ISSU DES SCIENCES DE L'ESPRIT CONTEMPORAINES

La théorie psychanalytique repose sur un ensemble d'assomptions fondamentales que Freud appelait une métapsychologie, et qui consistaient en des spéculations sur la nature du fonctionnement biologique du cerveau (Fulgencio, 2005). Ces spéculations étant rendues depuis longtemps obsolètes par l'avènement des neurosciences cognitives, elles doivent donc être remplacées par un modèle théorique fondamental tiré des avancées scientifiques contemporaines. Notre *explanans* consistera donc en l'élaboration de ce modèle métathéorique fondamental, à travers une synthèse des travaux les plus pertinents provenant des différentes sciences de l'esprit contemporaines. Dans ce chapitre, notre objectif sera de construire un échafaudage théorique qui va nous permettre (au chapitre 3) d'étayer et d'améliorer substantiellement l'explication initiale de la psychothérapie psychodynamique présentée au chapitre précédent.

En suivant le cadre conceptuel et épistémologique de la théorie générale des systèmes et de la nouvelle philosophie mécaniste, nous faisons l'hypothèse que les mécanismes qui sous-tendent les phénomènes psychologiques impliqués dans la psychothérapie psychodynamique sont distribués sur une multitude d'échelles spatiotemporelles différentes. C'est pourquoi le cadre théorique explicatif élaboré dans ce chapitre sera organisé comme un échafaudage hiérarchique dans lequel chaque modèle théorique avancé aura pour fonction de couvrir un niveau d'organisation particulier des systèmes

cognitifs humains pertinents à l'explication des phénomènes en jeu au sein de la psychothérapie psychodynamique. En partant de l'organisation fondamentale des systèmes vivants, nous progresserons dans les échelles afin de couvrir la dimension neurologique, écologique, puis socioculturelle de la cognition humaine.

2.1 L'organisation fondamentale des systèmes vivants

Le principe de l'énergie libre (*Free Energy Principle*; FEP) est le modèle théorique fondamental autour duquel va s'organiser l'ensemble de ce chapitre (Friston et al., 2006). Ce dernier part de la prémisse selon laquelle, afin de pouvoir formuler une théorie unifiée du fonctionnement des systèmes neurocognitifs humains, il nous faut tout d'abord comprendre les principes d'organisation fondamentaux des systèmes vivants dont ils émergent (Badcock *et al.*, 2019 ; Ramstead *et al.*, 2017). En effet, et tel que la théorie générale des systèmes nous l'enseigne, les processus qui définissent l'organisation des systèmes neurocognitifs humains doivent nécessairement obéir aux contraintes des principes organisationnels des systèmes biologiques et physiques dont ils émergent (Connolly et van Deventer, 2017). Si l'on suit cette ligne de pensée, alors il est logique de considérer que les théories élaborées au sein des neurosciences cognitives (et des sciences de l'esprit en général) devraient être contraintes par, c'est-à-dire formulées de manière à être compatibles avec, les connaissances en biologie et en physique. Conformément à cette idée, le FEP consiste en une tentative de déterminer le principe fondamental qui définit l'organisation de ces systèmes physiques particuliers que sont les systèmes vivants, et ce dans le but de l'utiliser afin de fournir une base métathéorique universelle permettant de réguler et de coordonner les théories qui sont développées au sein des sciences de l'esprit contemporaines (Friston, 2010).

Cette contrainte théorique commune aurait pour avantage de faciliter l'unification des différents champs d'études qui composent les sciences de l'esprit contemporaines, en favorisant l'intégration des modèles théoriques qu'elles développent, puisque ces derniers seraient organisés autour d'un même principe fondamental. Dans cette première section de notre deuxième chapitre, nous allons procéder à l'exposé de trois modèles théorique sur l'organisation fondamentale des systèmes vivants qui viendront contraindre la compréhension du fonctionnement cognitif humain que nous développerons par la suite : la théorie de l'autopoïèse (Maturana et Varela, 1980), le principe de l'énergie libre (Friston, 2010), et la neuroéthologie variationnelle (Ramstead *et al.*, 2017).

2.1.1 L'autopoïèse

Dans leur ouvrage de 1980 intitulé *Autopoiesis and Cognition : the Realization of the Living*, Humberto Maturana et Francisco Varela proposent ce qu'ils considèrent comme une définition nécessaire et suffisante de ce qu'est un être vivant (Maturana et Varela, 1980). En partant de la théorie des systèmes, ils posent la prémisse selon laquelle l'unité de base de la vie au sens biologique du terme, l'organisme, doit être comprise comme une catégorie particulière de système physique : les systèmes vivants. Ils associent cela à une approche mécaniste qui stipule que l'explication d'un phénomène doit montrer comment celui-ci est généré à partir des interactions entre les éléments qui le composent. La combinaison de ces deux perspectives les amène à considérer un système vivant comme une machine : un système qui est fait de relations entre les éléments qui le composent, et qui tire son unité des processus mis en place à travers ces relations. Une machine (ou un système) peut-être décrite selon deux aspects fondamentaux : son organisation et sa structure. L'organisation d'une machine

correspond à l'ensemble des relations qui définissent ce système en tant qu'unité, et déterminent les dynamiques des interactions et des transformations qu'elle peut subir sans perdre cette unité. La structure d'une machine correspond quant à elle à la manière spécifique dont les relations qui s'opèrent entre les éléments s'instancient en une machine concrète.

Maturana et Varela affirment que le développement d'une véritable compréhension de ces machines particulières que sont les systèmes vivants doit passer par une explicitation de leur fonctionnement, de telle manière que les différentes formes structurelles (individuelles) qu'elles peuvent prendre apparaissent comme une conséquence de leur organisation (commune). C'est pour accomplir cette tâche qu'ils développent le concept *d'autopoïèse*, à travers lequel ils entendent spécifier le type d'organisation particulière qui caractérise l'ensemble des systèmes vivants. Le néologisme d'autopoïèse vient du grec, le préfixe *auto* signifiant « soi-même », et le suffixe *poiësis* signifiant « production » ou « création ». Il y a deux conditions organisationnelles pour qu'un système soit considéré comme autopoïétique :

1. L'autoproduction : une machine autopoïétique génère et spécifie sa propre organisation à travers son opération en tant que système de production de ses propres composantes. En effet, l'organisation d'une machine autopoïétique correspond à un réseau de processus de production des éléments qui composent la machine elle-même. À travers leurs interactions et leurs transformations, ces éléments régénèrent et réalisent continuellement le réseau de processus (relations) qui les produisent eux-mêmes. Ainsi, les systèmes vivants sont des machines autocréatrices : elles transforment la matière en « elles-mêmes » de telle manière que le produit de leur opération est leur propre organisation.

2. L'autodistinction : une machine autopoïétique, de par la cohérence interne des processus qui définissent son organisation, se maintient activement en tant qu'unité et

se différencie ainsi de son environnement physique. Les éléments qui la composent constituent la machine en tant qu'unité concrète, et ce en spécifiant le domaine topologique de sa réalisation en tant que réseau. Autrement dit, un système vivant possède une limite claire qui le différencie de son environnement. Cette condition est capturée par le concept de *clôture opérationnelle*, qui exprime la manière dont un organisme affirme son identité en installant une séparation claire entre ses états internes et les états externes qui appartiennent à son environnement (Kirchhoff *et al.*, 2018). Le terme de clôture ne veut pas dire qu'un système est matériellement ou énergétiquement fermé au monde externe (ce qui est impossible). Au contraire, un système autonome est un système qui se maintient loin de l'équilibre thermodynamique⁴, et qui échange continuellement de la matière et de l'énergie avec son environnement.

En effet, le processus autopoïétique qui définit un système vivant dépend des ressources matérielles et énergétiques qu'il obtient dans son interaction avec son milieu externe, ce qui implique que son organisation doit nécessairement être accordée intimement avec les dynamiques organisationnelles de ce milieu. Un autre concept important de la théorie autopoïétique du vivant est celui de couplage structurel, qui réfère à l'histoire des interactions récurrentes entre deux systèmes ou plus, et qui mène à une congruence structurelle entre eux. Selon la conception autopoïétique du vivant, les changements d'état que subit un système autonome résultent de sa clôture opérationnelle, et de son couplage structurel : le système se différencie de son environnement tout en restant fonctionnellement imbriqué dans, et dépendant des interactions avec ce dernier. Le résultat de n'importe quel changement d'état est

⁴ En physique thermodynamique, on dit d'un système fermé qu'il a atteint l'équilibre thermodynamique lorsqu'il n'évolue plus, c'est-à-dire que ses variables d'état (vitesse, température, énergie) ne changent plus à travers le temps. L'équilibre thermodynamique n'est possible que pour un système fermé, puisque toute influence de l'extérieur sur le système va mener à une modification des variables d'état du système.

toujours d'amener plus d'activité auto-organisée au sein du système, du moins si sa clôture n'est pas interrompue. Autrement dit, puisqu'un système est caractérisé par son organisation, tout changement durable de l'état d'un système correspond à une modification de son organisation. Pour continuer à exister à travers le temps, un système autopoïétique doit faire coïncider son organisation avec les conditions offertes par son milieu externe, de manière à ce que le processus d'autoproduction qui permet le maintien de son organisation et de sa clôture opérationnelle puisse se renouveler. C'est ce phénomène de couplage structurel entre les organismes et leurs environnements que le célèbre concept biologique *d'adaptation* tente de capturer. En ce qui nous concerne, nous allons tenter d'utiliser le concept de couplage structurel pour expliciter la manière dont l'organisation du système nerveux (et cognitif) d'un individu humain est déterminée par les dynamiques d'interactions récurrentes avec son environnement social proximal durant l'ontogénie et au-delà.

La théorie de l'autopoïèse constitue l'une des premières tentatives modernes de comprendre l'organisation fondamentale des systèmes vivants afin d'informer et de contraindre l'étude de la cognition de ces systèmes vivants particuliers que sont les êtres humains. Ce projet est basé sur la thèse de la continuité entre la vie et l'esprit, qui stipule que les propriétés organisationnelles complexes de la cognition humaine sont en réalité des manifestations de principes fondamentaux de l'organisation fondamentale des systèmes vivants en général (De Jesus, 2016). Une manière imagée de présenter cette thèse serait d'affirmer que dans chaque système vivant il existe une forme prototypique d'esprit, et qui correspond aux habiletés (proto-)cognitives minimales lui permettant de percevoir et d'interagir efficacement avec son environnement de manière à maintenir l'organisation qui le définit (Thompson, 2007). À bien des égards, le principe de l'énergie libre peut être considéré comme une version améliorée de la théorie de l'autopoïèse, puisqu'elle récupère, reformule, et étend ses principes à l'aide des outils conceptuels et formels de la physique thermodynamique et des sciences computationnelles afin de les opérationnaliser dans le cadre des

neurosciences cognitives contemporaines (Kirchhoff, 2018). C'est vers cette théorie contemporaine sur l'organisation fondamentale des systèmes vivants que nous nous tournons maintenant.

2.1.2 Le principe de l'énergie libre

Au tournant du millénaire, une nouvelle théorie en neurosciences appelée le principe de l'énergie libre (*Free Energy Principle*; FEP) a vu le jour (Friston et Klaas E., 2007). Puisé principalement de la thermodynamique statistique et du *machine learning*, le FEP est un modèle formel des processus neuronaux qui a été initialement proposé pour expliquer la perception, l'apprentissage et l'action, mais qui a depuis été étendu pour expliquer l'évolution, le développement, la forme, et la fonction du cerveau. Plus récemment, il a aussi été appliqué aux systèmes biologiques à travers les échelles spatiales et temporelles, allant de phénomènes à une échelle microscopique (auto-organisation dendritique et morphogénèse), en passant par les échelles intermédiaires (ensembles culturels), jusqu'aux échelles macroscopiques (la sélection naturelle). Afin de fournir une introduction satisfaisante à ce modèle théorique complexe que nous ne pouvons absolument pas prétendre maîtriser dans toute sa profondeur, nous allons nous appuyer directement sur l'explication qui en est faite par ses créateurs dans un article récent (Ramstead *et al.*, 2017).

En 1944, le célèbre physicien Erwin Schrodinger posait la question suivante : qu'est-ce que la vie ? (Schrodinger, 2012). Il affirmait alors que parmi les systèmes physiques, les systèmes vivants sont uniques en ce qu'ils semblent, en apparence, résister à la seconde loi de la thermodynamique. La seconde loi de la thermodynamique stipule que tous les systèmes physiques doivent inexorablement tendre vers l'entropie (le désordre). Or, les systèmes vivants semblent parvenir à maintenir leur structure et leur organisation à travers le temps. Pour expliquer les systèmes vivants, le FEP part d'une prémisse très similaire à celles qui sont au fondement de la théorie de l'autopoïèse : tous les systèmes biologiques démontrent une forme spécifique d'auto-organisation.

Celle-ci a été sculptée par la sélection naturelle pour leur permettre de maintenir activement leur intégrité en revisitant des états caractéristiques, à l'intérieur des limites bien définies de leur espace des phases. Selon la théorie des systèmes dynamiques, l'espace des phases est un espace abstrait dont les coordonnées correspondent aux variables du système étudié (Terman et Izhikevich, 2008), où l'on peut donc représenter tous les états possibles du système, chaque état possible correspondant à un point unique dans l'espace des phases. En d'autres termes, le FEP stipule qu'il y a une forte probabilité qu'un organisme occupe un ensemble d'états limité, relativement petit — son ensemble de viabilité — dans l'ensemble total des états possibles qu'il peut occuper (son espace des phases). Dans les termes de la théorie de l'information, cela veut dire que la fonction de densité de probabilité qui décrit les états possibles du système a une faible entropie.

Cela est un fait remarquable à propos des systèmes vivants. Alors que tous les autres systèmes physiques auto-organisés, des flocons de neige aux systèmes solaires, suivent un chemin inévitable et irréversible vers le désordre (l'entropie) ; les systèmes vivants revisitent constamment un ensemble d'états limités. Il s'ensuit que les systèmes biologiques sont caractérisés par un attracteur dynamique aléatoire — un ensemble d'états attractifs qui sont fréquemment revisités (Milnor, 2006). En effet, les caractéristiques à travers lesquelles nous définissons les systèmes vivants sont simplement des affirmations à propos des états attractifs caractéristiques dans lesquels nous les trouvons. Cet ensemble d'états attractifs peut être interprété comme le phénotype étendu de l'organisme — sa morphologie, sa physiologie, ses régularités comportementales, ses régularités culturelles, et même les environnements construits qu'il aménage. Un exemple simple est celui du poisson qui est contraint de par son phénotype de se maintenir dans l'eau sous peine d'une mort certaine. Un exemple plus complexe est celui du castor, qui se nourrit de bois et utilise aussi cette ressource pour construire des barrages afin d'y habiter et de s'y reproduire.

Ainsi les systèmes vivants se maintiennent activement dans un ensemble d'états qui les caractérisent, ce faisant ils réduisent au maximum l'entropie qui les habite, en empêchant leurs états de se dissiper sur leur espace des phases. Ce comportement différencie les états de vie par rapport aux autres états, comme la dégradation ou la mort. En termes informationnels, les systèmes vivants peuvent être décrits comme travaillant constamment pour maximiser la probabilité de se trouver dans le même ensemble d'états limité. Comme cet ensemble d'états limité (leur phénotype étendu) est conçu comme ce qui les caractérise en tant que système vivant, on dit qu'ils sont « autoprouvants » (*self-evidencing*; Hohwy, 2016) : ils se déplacent (dans leur espace des phases) pour maximiser les preuves de leur existence. La question devient alors : comment réalisent-ils cela ?

C'est ici que le principe de l'énergie libre (FEP) entre en jeu. Il affirme que tous les systèmes biologiques maintiennent leur intégrité en réduisant activement le désordre ou la dispersion (entropie) de leurs états sensoriels et physiologiques, et qu'ils font cela en minimisant leur *énergie libre variationnelle*. L'énergie libre variationnelle est une information théorique fonctionnelle qui se distingue de l'énergie libre thermodynamique (Helmholtz), mais dont les caractéristiques sont très similaires. Techniquement, l'énergie libre variationnelle est une quantité théorique informationnelle qui limite (en étant plus grande que) l'entropie des échanges sensoriels entre un système biologique et son environnement. Puisque le répertoire d'états fonctionnels ou adaptatifs occupés par un organisme est limité, la distribution des probabilités de ces états caractéristiques à une faible entropie : il y a une forte probabilité pour que l'organisme revisite un petit nombre d'états. Ainsi, l'impératif distal de survie d'un organisme, et le maintien de ses états fonctionnels dans les limites de sa physiologie (homéostasie et allostasie) se traduisent dans un évitement proximal de la *surprise*. Ici, le concept de surprise est à comprendre, non pas dans un sens psychologique, mais dans un sens statistique, qui stipule qu'un état surprenant correspond à un état qui a une probabilité relativement faible de se produire. Un état

surprenant ne fait donc pas partie de l'ensemble de viabilité de l'organisme, qui est composé des états qui ont la plus forte probabilité de se produire. Cette propension à minimiser la surprise est le résultat de la sélection naturelle : les systèmes auto-organisés qui sont capables d'éviter une entropie interne trop élevée ont été sélectionnés par rapport à ceux qui ne pouvaient pas. Il nous reste à répondre à la question suivante : comment les systèmes biologiques parviennent-ils à minimiser leur énergie libre ?

Pour comprendre la différence, subtile, mais importante, entre surprise et énergie libre, nous devons regarder avec plus d'attention ce qui constitue un système. Tout d'abord, nous devons différencier le système de son environnement, ces états qui constituent ou sont intrinsèques au système (les états internes) et ceux qui ne le sont pas (les états externes). C'est pour cela que l'on doit introduire le concept de « couverture de Markov », qui formalise la notion de clôture opérationnelle des systèmes vivants (Kirchhoff et al. 2018). Une couverture de Markov est donc une troisième classe d'états qui opère la médiation entre les états internes et externes. Les couvertures de Markov établissent une *indépendance conditionnelle* entre les états internes et externes et rendent l'intérieur ouvert à l'extérieur, mais seulement sous certaines conditions (les états internes ne « voient » seulement les états externes qu'à travers la « lentille » de la couverture de Markov). La couverture de Markov peut ensuite être divisée entre des états « sensoriels » et « actifs » qui sont distingués de la manière suivante : les états internes ne peuvent pas influencer les états sensoriels, les états externes ne peuvent pas influencer les états actifs. Avec ces indépendances conditionnelles en place, nous avons maintenant une séparation (probabiliste) bien définie entre les états internes et externes d'un système. À titre d'exemple, nous pouvons penser ici à la membrane d'une cellule, ou la surface épidermique du corps humain (incluant les organes perceptuels).

Avec les couvertures de Markov en tête, il est raisonnablement aisé de montrer que les états internes doivent — en vertu de la minimisation de la surprise — encoder une

distribution de probabilités des états externes ; c'est-à-dire, des causes des impressions sensorielles sur la couverture de Markov. Cela nous ramène à l'énergie libre. L'énergie libre est une fonctionnelle (la fonction d'une fonction) qui décrit la distribution de probabilité encodée par les états internes d'une couverture de Markov. Notons que cela est différent de la surprise, qui est une fonction de la probabilité des états de la couverture de Markov elle-même. En d'autres termes, l'énergie libre est une fonction de croyances probabilistes, encodées par les états internes à propos des états externes (attentes à propos des causes probables des entrées sensorielles). Quand ces croyances sont égales à la probabilité postérieure des états externes, l'énergie libre devient équivalente à la surprise. Autrement, elle est toujours un petit peu plus grande que (elle impose une limite supérieure à) la surprise. Cela veut dire que les systèmes vivants peuvent être caractérisés comme minimisant l'énergie libre variationnelle, et donc la surprise, où la minimisation de l'énergie libre variationnelle implique l'optimisation des croyances à propos des choses au-delà ou derrière la couverture de Markov (le monde extérieur).

L'introduction à la notion de couverture de Markov nous permet de préciser le FEP. Le FEP nous dit comment les quantités qui définissent les couvertures de Markov changent lorsque le système se déplace vers son minimum d'énergie libre variationnelle. Il affirme que pour qu'un système résiste à l'érosion entropique et se maintienne dans un ensemble d'états limité (pour posséder une homéostasie généralisée) il doit instancier un modèle statistique et causal de sa relation avec son éconiche. En d'autres termes, un organisme ne fait pas juste encoder un modèle du monde, il est un modèle du monde — une transcription physique des régularités causales dans son éconiche qui a été sculptée par des interactions réciproques (couplage structurel) entre son auto-organisation et la sélection naturelle (pressions environnementales) à travers le temps.

En ce qui nous concerne, le FEP va constituer le fondement du cadre théorique explicatif que nous allons tenter d'élaborer pour expliquer les phénomènes en jeu au sein de la psychothérapie psychodynamique. Ainsi, l'ensemble des modèles théoriques présentés dans la suite de ce chapitre, et qui seront utilisés au chapitre 3 dans notre tentative d'explication, devront être contraints par le FEP, c'est-à-dire qu'ils devront être réductibles aux — explicable en termes des — dynamiques de minimisation de l'énergie libre qui s'imposent à l'ensemble des systèmes vivants. Dans la section qui suit, nous nous tournons vers une tentative récente d'utiliser le FEP afin d'élaborer une ontologie formelle des systèmes vivants qui permettrait l'intégration des modèles théoriques hétérogènes des différentes sciences de l'esprit contemporaines au sein d'un cadre métathéorique unifié.

2.1.3 La neuroéthologie variationnelle

En combinant le cadre théorique du FEP, et en particulier le formalisme de la couverture de Markov, avec le cadre conceptuel de la théorie générale des systèmes, Ramstead et al. (2017) ont proposé une ontologie formelle des systèmes vivants : la neuroéthologie variationnelle. Le point central de leur argument est que les organismes peuvent être décrits en termes d'un espace des phases (hautement dimensionnel) induit par des couvertures de Markov hiérarchiquement imbriquées qui occupent des régions hiérarchiquement nichées dans l'espace des phases total des systèmes vivants. Pour se représenter de telles dynamiques hiérarchiques, il est utile d'introduire la notion d'un espace d'échelle. Les espaces d'échelle nous permettent d'observer des structures à différentes échelles spatiales. Imaginez que vous prenez une photographie, et que vous zoomez progressivement pour examiner de plus petits détails. Lorsque vous zoomez, vous traversez un espace d'échelle (spatial). La notion d'un espace d'échelle est utile

parce que l'augmentation dans l'échelle (spatiale), lorsque nous nous déplaçons (vers le haut) d'un niveau hiérarchique de couvertures de Markov à l'autre, implique nécessairement une augmentation de l'échelle temporelle. Notons que la notion d'espace d'échelle a un usage purement descriptif. Le système sous-jacent en question ne change pas — juste son niveau de description, la manière dont il est mesuré, ou la perspective prise sur son auto-organisation hiérarchique.

Les auteurs poursuivent en affirmant qu'ainsi : nous sommes en mesure de décrire l'univers en termes de couvertures de Markov imbriquées dans des couvertures de Markov, allant des échelles microscopiques jusqu'aux échelles macroscopiques. Pour expliciter cela, considérons un ensemble de cellules, chacune équipée d'une couverture de Markov qui correspond à la surface de la cellule. Parce que les états internes de chaque cellule sont séquestrés derrière leur couvertures de Markov respectives, toutes les interactions entre les cellules doivent être médiées par leurs couvertures de Markov. Cela veut dire que nous pouvons décrire l'auto-organisation de l'ensemble de cellules purement en termes de transactions entre les états (sensoriels et actifs) des couvertures de Markov. Cependant, ces échanges vont eux-mêmes avoir une structure éparpillée qui induit une couverture de Markov de couvertures de Markov. Par exemple, un sous-ensemble de l'ensemble pourrait être un organe qui est entouré d'une membrane épithéliale qui fournit une couverture de Markov pour toutes les cellules qui constituent les états internes de l'organe. Cependant, cela suit toujours exactement la même structure (probabiliste) — un ensemble de couvertures de Markov. Nous pouvons donc répéter ce processus à des échelles d'auto-organisation de plus en plus grandes, pour créer une série de systèmes hiérarchiquement nichés (le corps par exemple).

Cette sorte de structure hiérarchique fournit une perspective universelle et récursive pour comprendre l'auto-organisation à travers les échelles spatiales et temporelles, et pour expliquer comment chaque niveau contextualise (contraint) les niveaux autant au-dessus qu'en dessous. Le point clé ici est qu'à tout niveau, les mêmes dynamiques

variationnelles de réduction de la surprise doivent être en jeu pour fournir des couvertures de Markov pour le niveau supérieur.

La formulation hiérarchique de l'énergie libre présentée ici englobe les façons dont le FEP s'applique à chaque niveau d'organisation hiérarchique en articulant la minimisation de l'énergie libre à travers les échelles. N'importe quelle opération ou processus du système (un évènement particulier étendu dans le temps dans cette ontologie) peut être vu comme se déployant à travers des échelles spatiales, poussé par des dynamiques temporellement intégrées de minimisation de l'énergie libre. Sur un laps de temps donné, la minimisation de l'énergie libre peut donc être saisie comme un déploiement hiérarchique, à travers différentes échelles spatiales, d'une même dynamique invariante (indépendante de l'échelle). Puisque les échelles temporelles et spatiales sont intrinsèquement corrélées (les évènements qui se déploient sur de longues distances prennent habituellement plus de temps pour se déployer), ce que nous avons est un espace d'échelle qui est peuplé majoritairement le long de sa diagonale (spatiotemporelle).

Formuler les choses ainsi nous permet de dériver une typologie dynamique, opérationnelle sur le plan computationnel, pour les évènements scientifiquement pertinents (types de systèmes biotiques) à chaque échelle temporelle, et aussi de distinguer entre différentes régions de l'espace des phases biophysique, qui correspond grossièrement aux ontologies des disciplines scientifiques concernées par différents systèmes vivants (biologie, psychologie, et sciences sociales). Ces impératifs variationnels auto-organisés restent identiques à chaque niveau d'organisation, mais diffèrent fondamentalement dans la manière dont ils sont médiés.

La notion des couvertures de Markov nous permet de définir une ontologie formelle des systèmes vivants. Des couvertures de Markov sont nichées dans des couvertures de Markov, à des échelles spatiales et temporelles différentes. Cela suggère qu'il doit y

avoir une consistance interne (indépendante de l'échelle), dans le sens que l'énergie libre variationnelle associée avec une couverture de Markov globale (espèces) doit être conforme aux mêmes principes que toutes les couvertures de Markov qui la constituent (phénotypes), ce qui s'applique récursivement jusqu'au niveau des macromolécules biologiques. L'avantage de cela est que cela nous permet de modéliser de manière computationnelle la dynamique de minimisation de l'énergie libre à travers les échelles temporelles et spatiales. À mesure que nous montons les couvertures de Markov nichées, et que nous considérons différents systèmes biologiques, l'énergie libre est minimisée de manières dynamiquement équivalentes, mais mécaniquement hétérogènes.

Cela nous amène à une ontologie métathéorique des systèmes biologiques dérivée du FEP — la neuroéthologie variationnelle — qui peut être utilisée pour expliquer et prédire comment les systèmes vivants, à n'importe quelle échelle temporelle ou spatiale, instancient la dynamique de minimisation de l'énergie libre. La nature précise des mécanismes de minimisation de l'énergie libre va varier d'organisme en organisme, d'espèce en espèce, en fonction de leurs phénotypes et de leurs régularités biocomportementales. En d'autres termes, le FEP fournit un principe universel qui s'étend à travers toutes les échelles spatiales et temporelles, mais les manières particulières grâce auxquelles il est implémenté vont varier selon les espèces, les organismes, et échelles qui nous intéressent.

La neuroéthologie variationnelle nous offre une ontologie formelle des systèmes vivants qui est en directe continuité avec l'épistémologie générale des systèmes, qui est compatible avec notre conception mécaniste de l'explication en science, et qui a été spécifiquement conçue pour les neurosciences cognitives contemporaines. Nous pensons que ce modèle ontologique constitue une piste de solution (partielle) intéressante au problème de l'unification des sciences de l'esprit contemporaine. En effet, selon la perspective de la neuroéthologie variationnelle, les différentes sciences

de l'esprit pourraient être définies et organisées hiérarchiquement en fonction des espaces d'échelles du fonctionnement neurocognitif humain qu'elles étudient. Le formalisme des couvertures de Markov nous permet de considérer les phénomènes et les mécanismes étudiés par chaque discipline comme les manifestations de niveaux d'organisation spécifiques qui sont imbriqués dans un échafaudage hiérarchique et récursif. Ici, l'émergence d'une nouvelle échelle d'organisation correspond à l'émergence d'une nouvelle couverture de Markov qui vient étendre l'organisation du système à une nouvelle échelle spatiotemporelle. Puisque toute nouvelle échelle d'organisation doit nécessairement respecter les principes d'organisation de l'échelle dont elle émerge, la même relation de contrainte doit s'appliquer entre les modèles théoriques développés par les disciplines qui étudient ces échelles. Cela semble évident : les théories en biologie doivent respecter les lois de la physique.

Cependant, les espaces d'échelles adressés par les différentes sciences de l'esprit présentent un niveau de recouvrement important, et plusieurs d'entre elles étudient des phénomènes dont les mécanismes sont probablement distribués sur des échelles spatiotemporelles similaires, mais selon des méthodologies différentes. Par exemple, la psychologie cognitive traditionnelle étudie la cognition en mesurant la performance dans des tâches cognitives, alors que la neuropsychologie cherche à identifier l'activation neuronale qui émerge lorsque ces tâches sont effectuées. Bien sûr, ces méthodes ont déjà fait l'objet de tentatives d'intégration et de combinaison extensive, donnant lieu à l'avènement de ce que l'on appelle aujourd'hui le champ d'études général des neurosciences cognitives. Présenté ainsi, il est possible de considérer qu'en dehors du fait de proposer la compatibilité avec le FEP comme contrainte générale à l'élaboration de modèle théorique dans l'ensemble des sciences cognitives, la neuroéthologie variationnelle n'apporte finalement pas grand-chose de nouveau en ce qui concerne l'intégration interdisciplinaire dans les sciences de l'esprit. Cependant, cela n'est vrai que si l'on se limite aux domaines de phénomènes comme la perception

par exemple, dont l'étude est déjà bien avancée dans le cadre théorique dominant des neurosciences cognitives.

En effet, les dernières décennies ont vu naître une multitude de courants théoriques dont le but est de combler les angles morts des neurosciences cognitives contemporaines en mettant l'accent sur l'étude de phénomènes comme la conscience ou l'influence de la culture sur la cognition qui avaient longtemps été laissés de côté. C'est ainsi que sont apparues des sous-disciplines comme les neurosciences affectives (Panksepp, 2004) et les neurosciences sociales (Cacioppo *et al.*, 2002) par exemple, cherchant à adresser respectivement l'importance des processus émotionnels et le rôle déterminant de l'environnement social dans le fonctionnement cognitif humain. Ces courants théoriques adressent spécifiquement les phénomènes en jeu dans la psychothérapie psychodynamique, et qui sont au cœur de la conception de l'esprit entretenue par la psychologie psychodynamique. L'émergence conjointe de ces courants théoriques et du cadre ontologique de la neuroéthologie variationnelle permettant d'intégrer leurs modèles à l'ensemble des neurosciences cognitives constitue selon nous le contexte idéal afin de formuler un nouveau cadre explicatif pour la psychothérapie psychodynamique. Les outils conceptuels offerts par la neuroéthologie variationnelle vont nous permettre de préciser notre projet en formulant un nouvel objectif pour notre travail d'explication : définir l'espace d'échelle qui correspond aux phénomènes en jeu au sein de la pratique de la psychothérapie psychodynamique.

2.2 Application du principe de l'énergie libre au cerveau humain : une théorie computationnelle du fonctionnement neurologique

2.2.1 Les approches du traitement prédictif

L'application du principe de l'énergie libre en neurosciences computationnelles prend majoritairement la forme d'un ensemble de modèles qui considèrent le cerveau comme une machine à inférence : les approches du traitement prédictif. Notre tentative d'explication de la psychothérapie psychodynamique va être en grande majorité formulée dans le langage théorique proposé par ces approches. Selon elles, le cerveau correspond à un système dynamique complexe dont la principale fonction de minimiser l'énergie libre de l'organisme dont il fait partie en inférant les causes proximales et distales des stimulations sensorielles (états sensoriels des couvertures de Markov) qu'il reçoit (Hohwy, 2013). La stratégie employée par le cerveau, selon ces approches, est d'utiliser un modèle génératif des causes distales de ses états sensoriels, et ce afin de les prédire, et de maintenir l'organisme dans son ensemble de viabilité. La fonction du système est donc de prédire les états sensoriels à venir (prédictions) et de les comparer aux états sensoriels actuels (erreurs de prédiction), tout en minimisant la différence entre ces deux distributions (Ramstead *et al.*, 2017). Un modèle génératif correspond à une modélisation probabiliste de causes dans l'environnement, réalisée à partir des conséquences observées (données sensorielles), et dans le but de les expliquer afin de mieux les prédire.

Les erreurs de prédiction peuvent être minimisées de deux manières. Le système peut mettre à jour les modèles génératifs, changeant le modèle pour correspondre au monde, ce qu'on appelle *l'inférence perceptuelle*. Le système peut aussi performer des actions pour amener les états sensoriels en ligne avec les prédictions, changeant le monde pour correspondre au modèle, ce qu'on appelle *l'inférence active*. Ainsi, l'inférence

perceptuelle correspond à une modification de la manière dont les états internes de la couverture de Markov reflètent les régularités causales qui structurent l'environnement, et l'inférence active vise à modifier directement cette structure afin de rétablir les conditions favorables au maintien de l'organisme dans les normes qui définissent son phénotype. Ces deux processus se déroulent continuellement et simultanément, ce qui implique une continuité profonde entre perception et action (Seth, 2013). C'est à travers ces cycles d'action et de perception que s'installe une forme de *circularité causale* dans laquelle les états internes et externes à la couverture de Markov de l'organisme s'influencent mutuellement (Kirchhoff, 2018).

Un aspect important de la conception du cerveau présentée ici est son organisation hiérarchique. La manière adéquate de comprendre la notion de *hiérarchie prédictive* est de concevoir l'organisme et le cerveau comme une multitude de couvertures de Markov imbriquées les unes dans les autres. Ainsi chaque couche de neurones dans le cerveau constitue un (sous-)système vivant à part entière qui tente de prédire adéquatement les inputs provenant des couches inférieures en minimisant sa propre énergie libre. C'est ainsi que l'on en vient à la formulation du cerveau comme incarnant une hiérarchie prédictive, où les niveaux supérieurs travaillent à minimiser l'erreur de prédiction en supprimant, c'est-à-dire en expliquant, l'activité des niveaux inférieurs. L'activité consistante avec les niveaux les plus élevés de la hiérarchie étant supprimée, seule se propage dans la hiérarchie l'erreur de prédiction, c'est-à-dire l'information qu'il reste à expliquer. L'organisation hiérarchique est importante, car elle permet de générer des attentes « top-down » (descendantes) à propos des échantillons sensoriels provenant du monde. Elle permet aussi de résoudre les ambiguïtés dans l'inférence et la représentation des causes des sensations extéroceptives et intéroceptives. Cela implique une progression dans la complexité des représentations : en allant du noyau thalamique et des cortex sensoriels primaires jusqu'aux cortex associatifs et paralimbiques, on va des sensations jusqu'aux concepts en passant par les perceptions.

Cette progression est reflétée dans l'étendue temporelle de ce qui est représenté (Carhart-Harris et Friston dans Fotopoulou et al., 2012).

Les erreurs de prédictions sont associées aux connexions neurales ascendantes qui véhiculent l'information vers les niveaux supérieurs de la hiérarchie cérébrale. Les prédictions sont quant à elles associées aux connexions neurales descendantes, majoritairement inhibitrices, qui suppriment l'activité des niveaux inférieurs de la hiérarchie cérébrale. Il est important de saisir que ces prédictions ne sont pas simples prévisions des entrées sensorielles, et ne s'appliquent pas seulement à l'activité perceptuelle du système nerveux, mais à l'ensemble de l'activité neurocognitive de l'organisme. Ainsi, une action (inférence active) performée par un organisme, comme cueillir un fruit pour le manger par exemple, est réalisée à travers l'implémentation d'un ensemble de signaux *top-down*, impliquant des prédictions motrices qui visent à modifier les entrées sensorielles afin de maintenir l'organisme dans son ensemble de viabilité (homéostasie). Les prédictions ou les erreurs de prédiction sont associées avec un certain degré de « précision » qui détermine leur influence sur le traitement hiérarchique subséquent. La pondération de la précision module l'étendue avec laquelle les erreurs de prédiction sont résolues en mettant à jour les modèles prédictifs (inférence perceptuelle) ou en performant des actions (inférence active). Cela mène à une interprétation de *l'attention* comme l'optimisation de la pondération de la précision, équilibrant l'influence relative des erreurs de prédiction et des attentes précédentes sur l'inférence perceptuelle. La modulation du gain synaptique a été proposée comme le mécanisme neural qui sous-tend la régulation de la précision par l'attention (Brown et Friston, 2013).

Dans ce qui suit, nous allons poursuivre et approfondir l'explicitation du cadre théorique des approches prédictives du cerveau humain en les appliquant directement aux dimensions du fonctionnement neurocognitif humain qui sont pertinentes pour l'explication de la psychothérapie psychodynamique.

2.2.2 L'allostasie, l'intéroception, et la conscience affective

À partir de cette section, nous entrons dans le domaine des phénomènes qui, selon le cadre explicatif initial avancé au chapitre précédent, sous-tendent le fonctionnement de la psychothérapie psychodynamique. Plus précisément, le modèle théorique avancé dans cette section consiste en une reformulation, dans le cadre du FEP et des approches prédictives du cerveau humain, de la thèse neuropsychanalytique sur le rôle central de la conscience et des affects dans le fonctionnement neurocognitif humain.

Maintenant que nous avons brièvement introduit le cadre conceptuel des approches prédictives, essayons de poursuivre notre compréhension du rôle et de la fonction du cerveau. Nous savons que tous les organismes partagent le même impératif : maintenir leur organisation à travers le temps en minimisant leur entropie interne. Le développement d'un système nerveux (central) constitue seulement le chemin particulier emprunté par certains organismes multicellulaires pour répondre aux pressions de la sélection naturelle. Autrement dit, le cerveau, comme tout organe, n'est que le prolongement de l'organisme en ce qu'il sert la même fonction : participer à l'ensemble des processus qui réalisent et maintiennent l'organisation du système vivant auquel il appartient. Pour ce faire, le cerveau doit faire en sorte de maintenir le système vivant dans les normes que lui dicte son phénotype. Ce faisant, il a été proposé que la fonction principale de cet organe qu'est le cerveau puisse être définie comme la réalisation de *l'allostasie* (Barrett, 2016) : le processus qui mène à la stabilité (homéostasie) de l'organisme à travers des changements physiologiques ou comportementaux. Ainsi, chaque cerveau accomplit la même tâche principale : s'assurer que les systèmes physiologiques qui composent le corps d'un animal aient les ressources nécessaires pour fonctionner, et ce afin que l'animal puisse se développer, survivre, et se reproduire. La croissance (le développement), la survie et la reproduction (et donc la transmission de gènes) requièrent un apport continu de ressources

biologiques, métaboliques ou autres. Les ressources métaboliques (ou autres) sont requises pour planifier et exécuter les mouvements physiques nécessaires pour acquérir ces ressources à l'origine (et pour se protéger des menaces et des dangers potentiels). Le concept d'allostasie constitue une prolongation dynamique, c'est-à-dire étendue dans le temps, de la notion d'homéostasie. L'idée est que l'organisme, s'il veut que ses besoins soient remplis de manière constante, doit pouvoir prédire ses états à venir pour anticiper les mesures qu'il devra mettre en place afin de se maintenir dans les normes que lui dicte son phénotype.

Grâce au FEP et aux approches prédictives, nous avons une idée de comment ce processus d'autorégulation est implémenté par le cerveau : il élabore des modèles prédictifs à partir des entrées sensorielles qu'il reçoit. En tant que sous-système biologique à part entière, le cerveau possède sa propre couverture de Markov, dont les états externes sont composés à la fois du milieu interne et du milieu externe à l'organisme. Le traitement des entrées sensorielles qui nous informent sur le monde extérieur est appelé l'extéroception, ce qui comprend la vision, l'audition, le toucher et l'odorat. Le traitement des entrées sensorielles qui nous informent sur le milieu interne de l'organisme est appelé l'intéroception. L'intéroception a été définie comme l'ensemble des modalités informant l'organisme à propos de l'état homéostatique du corps, à la fois en relation avec les expériences qui trouvent leur origine dans l'organisme (conscience cardiaque, faim), ou en dehors du corps (goût, odeur, toucher affectif, douleur ; Fotopoulou & Tsakiris, 2017). Faisons remarquer immédiatement le recouvrement des modalités sensorielles qui composent l'extéroception et l'intéroception. Cela est dû au fait que ces deux aspects du traitement informationnel opéré par le cerveau sont intimement liés, et s'opèrent en collaboration.

Dans un article récent, Mark Solms et Karl Friston (2018) allient leurs perspectives respectives, celles de la Neuropsychanalyse et du FEP, afin d'élaborer une explication fonctionnaliste du phénomène de la conscience, qu'ils associent, en continuité avec les

neurosciences affectives, à la modalité de l'intéroception. Dans leur explication des mécanismes fondamentaux qui seraient à l'origine du fonctionnement et de l'architecture du système nerveux central des animaux vertébrés, Solms et Friston accordent une place centrale aux affects, qu'ils identifient à cet aspect de la cognition humaine que nous appelons « la conscience », et qui correspond au fait de posséder une expérience subjective. Leur conception des affects nous sera nécessaire pour développer une compréhension adéquate de la dimension émotionnelle de la cognition humaine, l'un des aspects centraux de la psychologie psychodynamique. De plus, une explication de la conscience issue de l'intégration de la Neuropsychanalyse avec le FEP devrait nous aider à développer une meilleure compréhension des processus mentaux impliqués dans la pratique de la psychothérapie. Plus précisément, nous pensons qu'elle pourrait nous permettre, par opposition, de redéfinir l'inconscient dynamique de la psychologie psychodynamique dans des termes compatibles avec les neurosciences cognitives contemporaines.

La position de Solms et Friston a pour point de départ l'affirmation suivante : l'ensemble d'états internes limités qui définissent le phénotype des systèmes vivants (leur *être*), possède (pour eux) une *valeur existentielle*. En effet, le maintien des paramètres homéostatiques correspondant aux besoins vitaux de l'organisme (ex. température corporelle, métabolisme du glucose, hydratation, etc.) autour de valeurs spécifiques (innées) est une tâche critique en ce qu'elle permet la perpétuation de son organisation, et donc, le maintien de son *existence* en tant que système vivant. Selon les approches prédictives, ces valeurs correspondent à des *croyances prédéterminées* (« prior beliefs ») ou des « préférences » incluses dans le modèle génératif incarné par le cerveau. En d'autres termes, un état préféré est un état qu'un système vivant s'attend à occuper, car cet état permet le maintien de son existence.

Solms et Friston proposent que la mesure — par un système auto-organisé — de sa propre énergie libre, considérée subjectivement, c'est-à-dire, à travers l'intéroception,

mène à l'émergence de ce que l'on appelle « l'affect ». En d'autres termes, l'affect est cet aspect intrinsèquement (auto-) évaluatif de l'être biologique, prédiqué sur la croyance que la survie est désirable (ce qui est le système de valeur qui sous-tend toutes les formes de vie). La déviation d'une valeur préférée, pour chacun des paramètres homéostatiques qui définissent le phénotype d'un organisme, est enregistrée comme un affect négatif, et le retour vers cette valeur préférée est enregistré comme l'affect positif, ces affects étant qualitativement spécifiques aux paramètres en question (ex. taux de glucose et faim versus satiété). L'arrivée à la valeur préférée (« satiété ») résout l'affect, ce qui implique que l'affect n'est plus généré à ce point (idéal). L'affectivité en général est donc à la fois négative et positive, et enregistre la demande de travail constante pour remplir les besoins de l'organisme. Ainsi, l'affect peut être conçu comme un mécanisme d'alarme endogène qui enregistre l'existence et la direction des déviations vis-à-vis des valeurs préférées pour chacun des paramètres qui définissent le phénotype d'un organisme.

Selon Solms et Friston, ce mécanisme d'auto-évaluation correspond, dans le cadre du traitement prédictif, au concept de précision. La précision peut être définie comme la fiabilité, ou la variance inverse, des fluctuations sensorielles (dans différentes modalités sensorielles) et elle correspond à un aspect important de l'inférence : la représentation de l'incertitude. On peut concevoir la précision comme la confiance accordée aux conséquences (prédites) d'une action (inférence active) ou à une source d'entrées sensorielles (erreurs de prédiction). De manière heuristique, on peut se représenter les erreurs de prédiction comme des nouvelles (journalistiques) et la précision de ces informations comme la fiabilité que l'on accorde à leurs sources. La précision contrôle l'influence de l'erreur de prédiction sur l'action et la perception. Physiologiquement, la précision est associée au gain post-synaptique des populations de neurones du cortex (cf. les cellules superficielles pyramidales ; Brown et al., 2013; Feldman & Friston, 2010 ; Friston et al., 2006). Dans ce sens, la précision a été associée

— à travers la minimisation de l'énergie libre — avec l'activation sélective ou la sélection attentionnelle (Clark, 2013 ; Hohwy, 2013 ; Kanai *et al.*, 2015).

La précision choisie quelles erreurs de prédictions sont sélectionnées et assimilées par les modèles prédictifs, et détermine ultimement comment nous représentons le monde, ainsi que nos actions sur celui-ci. Selon cette perspective, on dit que la précision active le traitement prédictif, car en l'absence de précision, les erreurs de prédictions vont être incapables d'induire une synthèse perceptuelle, un comportement, ou une réponse neuronale. En d'autres termes, sans précision, les erreurs de prédictions seraient séquestrées au point de leur formation dans les zones d'entrée sensorielles. De plus, la précision n'est pas une seule valeur, mais elle s'incarne sous différentes formes selon les domaines sensoriels : chaque sensation, et chaque abstraction hiérarchique doivent être équipées d'une précision qui doit être optimisée. Si les cerveaux sont des machines à inférence qui assimilent les données extéroceptives, proprioceptives et intéroceptives à travers la prédiction, alors leur précision respective est *à propos de* quelque chose. Dans le domaine extéroceptif, elle se manifeste par l'attention et l'atténuation sensorielles qui sélectionnent les aspects du monde extérieur sur lesquels diriger la perception et la cognition (Brown *et al.*, 2013 ; Feldman et Friston, 2010 ; Frith *et al.*, 2000). Dans le domaine proprioceptif, elle correspond à la sélection et à la réalisation de prédictions motrices associées avec l'action et la sélection de buts, ce qui peut être capturé à travers le concept de *motivation* (Cisek et Kalaska, 2010 ; Frank, 2005 ; Friston *et al.*, 2012, 2014 ; Moustafa *et al.*, 2008). Dans le domaine intéroceptif, elle détermine les *ressentis affectifs*, c'est-à-dire, la meilleure explication pour les signaux intéroceptifs qui ont été sélectionnés (Hohwy, 2013 ; Seth, 2013).

Ce sont ces ressentis subjectifs qui définissent ce que l'on appelle l'expérience consciente, qui dans ce contexte, est conçue comme émergeant du processus *d'inférence intéroceptive* (Barrett et Simmons, 2015 ; Palmer *et al.*, 2015 ; Seth, 2013, 2014) et reflète l'optimisation de la précision des erreurs de prédictions qui portent sur

le milieu interne de l'organisme. L'expérience consciente, que Solms et Friston appellent « l'incertitude ressentie » ou l'affect, correspond selon eux à un impératif existentiel : il s'agirait du véhicule à travers lequel le système évalue et maintient son intégrité structurelle et fonctionnelle. Dans ce contexte, selon la perspective subjective (intéroceptive) d'un système vivant, la prédiction d'une augmentation de l'énergie libre est simplement « mauvaise » — il s'agit en effet d'une crise existentielle — alors que la prédiction d'une baisse de l'énergie libre est simplement « bonne ». Ces changements sont reconnus, intéroceptivement inférés (ressentis), en termes de changements dans l'incertitude et des ajustements concomitants de la prédiction.

Essayons maintenant de voir comment la perspective de Solms et Friston sur la conscience et les affects s'implémente concrètement dans l'architecture neuroanatomique du cerveau humain. L'organisme infère ses états viscéraux dans le but de minimiser les déviations par rapport aux valeurs homéostatiques préférées (*prior beliefs*) qui définissent ses besoins vitaux (température corporelle, métabolisme du glucose, hydratation, etc.). Cette tâche est effectuée principalement par ce que Solms et Friston appellent des « détecteurs de besoins », des structures du tronc cérébral comme l'hypothalamus médian, les organes circumventriculaires, et le complexe parabrachial, qui traitent des erreurs de prédictions intéroceptives. Si les déviations des valeurs prédites dépassent un certain seuil inné correspondant au phénotype de l'organisme, les erreurs de prédictions intéroceptives remontent la hiérarchie prédictive incarnée par le cerveau à travers l'activité du système réticulé activateur ascendant (ou ARAS, pour *Ascending Reticular Activating System*). Ce dernier, auquel l'on réfère aussi à travers l'acronyme ERTAS (*extended reticulo-thalamic activating system*), correspond à un ensemble de circuits neuronaux qui relie le tronc cérébral supérieur et le mésencéphale au cortex cérébral à travers différents réseaux passant notamment à travers le thalamus et l'hypothalamus. L'ARAS a été systématiquement associé à la conscience — tant au niveau des états de conscience (éveil vs sommeil) que des aspects

qualitatifs de la conscience comme les affects — à la régulation homéostatique, et ainsi qu'à la direction des ressources attentionnelles (Kovalzon, 2016).

L'une des thèses principales de la position de Solms et Friston consiste à affirmer que l'activation du système réticulaire activateur ascendant est ressentie à travers la conscience affective, et de plus, que les qualia perceptuels (la conscience extéroceptive) dépendent et sont précédés de cette activation affective qui est étendue vers la perception et la cognition (« je me sens comme ça à propos de ça »). Cette affirmation est basée sur plusieurs découvertes importantes de la recherche en neuropsychologie : a) l'entière conscience (à la fois affective et perceptuelle/cognitive) est oblitérée par de petites lésions du système réticulaire activateur ascendant (Parvizi et Damasio, 2001) ; alors que c) des lésions relativement larges du cortex mènent seulement à la perte de certains aspects de la conscience perceptuelle/cognitive (Penfield et Jasper, 1954). L'ensemble de ces connaissances pointent vers l'idée d'une relation de dépendance du cortex à l'ARAS qui, incarnant l'impératif de régulation de l'allostase, dirige l'activité du système neurocognitif dans son ensemble.

Ainsi, les erreurs de prédiction intéroceptives, qui évaluent les fonctions vitales du milieu interne, activent la fonction excitatrice du système réticulé activateur ascendant lorsque les déviations des valeurs homéostatiques préférées excèdent un certain seuil. Selon Solms et Friston, ces déviations homéostatiques génèrent des vecteurs d'excitation qualitativement différents (et sélectionnés par la modulation de la précision) en relation aux différents besoins de l'organisme, qui sont ressentis comme des affects homéostatiques — faim, sommeil, froid, excitation sexuelle, etc. — aussi appelés *pulsions* (« drives » ; Pfaff, 1999). Ces affects ont une valence négative, car ils incarnent la prédiction d'une augmentation de l'énergie libre, et entraînent la modulation de la précision, c'est-à-dire de l'incertitude prédite, vis-à-vis des états internes de l'organisme. Initialement, ce changement dans la précision mène au déploiement de prédictions proprioceptives instinctuelles (cf. réflexes), qui décident de

la mise en place de comportements stéréotypés (inférences actives automatiques) afin de remplir le besoin détecté, de résoudre l'affect, et donc l'incertitude vis-à-vis de l'état de fait actuel.

L'étendue des perturbations homéostatiques qui peuvent être réglées par des réflexes autonomes est très limitée ; elles demandent habituellement à être résolues par la mise en place de mesures qui visent la réduction de l'énergie libre en tenant compte du contexte (extéroceptif) : des *actions intentionnelles*. Ainsi, la résolution de l'incertitude, et donc l'optimisation la précision, implique la sélectivité et sous-tend donc un choix. Pour que l'animal survive dans des environnements imprévisibles, les plans d'action instinctuels doivent subir une complexification graduelle qui s'opère à travers un processus d'apprentissage réalisé *à partir de l'expérience* (ou le vécu), de l'organisme. En effet, lorsque les inférences actives automatiques sont inefficaces et ne parviennent pas à minimiser l'erreur de prédiction, la persistance de l'incertitude ressentie (affects négatifs) exerce une demande de travail sur le système qui se caractérise par l'excitation du cortex préfrontal, la région du cortex associée aux fonctions exécutives et à l'attention, par le système réticulé activateur ascendant afin de générer et de mettre en place de nouveaux plans d'action. Principalement implémentée par des circuits dopaminergiques, cette excitation du cortex antérieur par l'ARAS donne lieu à l'émergence de ce qui a été conceptualisé comme un système polyvalent de recherche et d'exploration qui génère des comportements spontanés et des émotions « épistémophiles » comme la curiosité, l'optimisme, et l'enthousiasme, et met l'animal en contact avec des opportunités d'action lui permettant de satisfaire ses besoins (Wright et Panksepp, 2012). L'existence de ce système explique en partie pourquoi les animaux ne peuvent pas minimiser leur énergie libre en trouvant et en restant dans une pièce noire ou les conditions ne changent jamais.

La dynamique d'activité exploratoire décrite ci-dessus est sous-tendue par la modulation de la précision accordée aux erreurs de prédiction extéroceptives et

proprioceptives, ce qui se manifeste à l'échelle cognitive par la direction de l'attention perceptuelle et la sélection de but. Les actions qui résultent de cette activité demandent de la plasticité, et sont volontaire. Puisqu'elles impliquent de l'incertitude, et donc un choix, elles doivent être guidées dans l'ici et maintenant par des prédictions à propos des conséquences (bonnes ou mauvaises) qu'elles vont avoir sur l'évolution valeurs des paramètres homéostatiques, et donc sur l'intégrité de l'état interne du système. Autrement dit, les prédictions nouvellement générées doivent être sélectionnées en fonction de leur capacité à minimiser efficacement l'énergie libre, c'est-à-dire, à remplir efficacement les besoins de l'organisme. Selon Solms et Friston, cette sélection s'effectue à travers la génération *d'affects positifs*, qui incarnent un changement dans l'incertitude ressentie, et qui informent le système d'un mouvement de retour vers les valeurs homéostatiques préférées⁵. En résumé, *l'apprentissage à partir de l'expérience* serait rendu possible par le ressenti affectif, qui correspond à l'évaluation consciente, encodée comme un changement dans la précision, qui guide (sélectionne) les perceptions et les actions en accord avec les préférences homéostatiques (valeurs existentielles). Solms et Friston affirment que c'est ici que réside la contribution causale de la conscience (de l'expérience subjective). Il s'agit du moyen, sélectionné par l'évolution naturelle, qu'on les animaux vertébrés d'enregistrer subjectivement (intéroceptivement) les dangers existentiels reliés au maintien de l'organisation définie par leurs phénotypes.

⁵ Le mécanisme décrit ici n'est pas sans rappeler celui du conditionnement du behaviorisme.

2.2.3 Un modèle des émotions selon les approches du traitement prédictif

À partir du travail effectué jusqu'à maintenant, nous pouvons désormais progresser vers une échelle spatiotemporelle supérieure de l'organisation des systèmes neurocognitifs humains, celle des phénomènes émotionnels. Pour ce faire, nous allons nous appuyer sur la position de Lisa Feldman Barrett, une auteure phare au sein des neurosciences affectives, et qui a développé durant les vingt dernières années une approche prédictive des émotions qui les place au fondement de l'architecture cognitive du cerveau humain (Barrett, 2017).

Selon Barrett (2016), la compréhension du cerveau offerte par l'approche prédictive implique la constatation suivante : un cerveau implémente son modèle interne avec des « concepts » qui « catégorisent » les sensations pour leur donner un sens (se référer au tableau 2.1.). Ici, Barrett fait équivaloir « prédictions » et « concepts ». En effet, les prédictions sont des concepts en ce qu'elles catégorisent ou « expliquent » l'ensemble des entrées sensorielles qui arrivent dans un cerveau à un moment donné. Cette thèse sur la définition neurocognitive de la notion de concept est issue d'une distinction opérée par Barsalou (1999, 2009) qui différencie entre : 1) les concepts comme simulateurs ; et 2) les concepts comme simulations. Cette différenciation s'intègre dans le cadre d'une théorie des symboles perceptuels, qui considère la cognition (représentationnelle) comme étant basée sur la réactivation neuronale d'états modaux (états sensoriels ou perceptuels). Selon cette théorie, alors qu'un simulateur intègre l'information à travers les instances d'une catégorie donnée, les simulations correspondent à des conceptualisations spécifiques de cette catégorie.

Derrière la première perspective se cache l'idée que le l'objectif premier de l'apprentissage humain consiste à développer des simulateurs. Un simulateur donné peut alors produire une infinité de simulations d'un certain type, où chacune des

simulations correspond à une conceptualisation différente de ce même type. Alors qu'un simulateur représente une catégorie de manière générale, une simulation fournit une manière spécifique de penser à cette catégorie. Les construits de simulateurs et de simulation correspondent à deux niveaux, ou strates, différent(e)s de la notion de concept. Cependant, lorsque Barrett (2016) affirme que les concepts sont équivalents à des prédictions, elle réfère aux concepts comme simulations, c'est-à-dire comme des instances (des événements) neurocognitives particulières d'une catégorie. Il nous semble que cette notion de concept neurocognitif est bien adaptée pour représenter le travail computationnel opéré par une population de neurones qui opèrent en synchronie pour faire sens des entrées sensorielles.

Selon Barrett, un cerveau implémente son modèle interne du monde à travers des concepts, car cela est efficace d'un point de vue métabolique (Barrett, 2016). Ce que cela veut dire, c'est qu'en élaborant des concepts, le cerveau compresse l'information qu'il reçoit de ses entrées sensorielles. Il simplifie celle-ci en y reconnaissant des patterns dans les entrées sensorielles, et en les catégorisant comme correspondant à des instances particulières de concepts, c'est-à-dire de situations qu'il a rencontrées auparavant. Un concept correspond donc à la modélisation de causes distales dans l'environnement (interne ou externe au corps) qui expliquent et prédisent le mieux l'ensemble des entrées sensorielles qui arrivent au cerveau. Barrett avance l'hypothèse selon laquelle tout apprentissage est apprentissage de concepts, puisque le cerveau condense les patterns d'activation redondants en les transformant en des résumés multimodaux plus efficaces (et moins coûteux). Ces « synthèses » restent disponibles pour la prédiction des futures entrées sensorielles. Ainsi, les concepts sont construits comme des résumés multimodaux : ce sont des *abstractions*.

Tableau 2.1. Les sens philosophiques et neurocognitifs des notions de catégorie et de concept selon Barrett (2016).

	Langage philosophique	Langage du cerveau
Catégorie	Une population d'événements ou d'objets qui sont traités comme étant similaires parce qu'ils servent tous un but particulier dans un contexte donné.	Des ensembles d'entrées sensorielles qui sont considérées comme étant similaires, car elles sont comprises comme renvoyant aux mêmes causes distales provenant du milieu interne ou externe à l'organisme (intéroception ou extéroception).
Concept	La population de représentations qui correspondent à ces événements ou ces objets.	Un groupe de « patterns » d'activité, de prédictions, distribués à travers une population de neurones, et qui modélise les conditions du milieu interne et du monde extérieur.

Si Barrett adopte cette signification spécifique de la notion de concept, c'est dans le but de l'appliquer à la compréhension de ces événements neurocognitifs particuliers que sont les émotions. Selon Barrett (2016), les émotions sont construites de la même manière que les autres perceptions, c'est-à-dire, à travers les mêmes mécanismes neuroanatomiques de traitement de l'information dans le cerveau. De ce fait, les émotions sont des concepts, elles sont des événements neurocognitifs qui implémentent

un processus de synthèse (de catégorisation) de l'information contenue dans un ensemble d'entrées sensorielles multimodales, et qui incarnent un modèle explicatif (abstrait) de leurs causes distales afin de les prédire, et de minimiser les erreurs de prédictions qui leur sont liées, et l'énergie libre du système dans son ensemble. Pour comprendre la spécificité des concepts émotionnels, il nous faut tout d'abord les distinguer des concepts affectifs.

En continuité directe avec la position de Solms et Friston (2018), Barrett conçoit les affects comme les événements neurocognitifs qui informent le système nerveux central de l'état du milieu interne de l'organisme dans le but de réaliser l'allostase (Duncan et Barrett, 2007). Ainsi, elle conçoit les concepts affectifs comme les modèles prédictifs issus de l'intégration des entrées sensorielles provenant de l'ensemble des modalités intéroceptives (Barrett et Simmons, 2015). Étant donné que le but central du cerveau est de réaliser l'allostase, alors il s'ensuit que totalités des concepts générés par notre système neurocognitif doivent contribuer de prêt ou de loin à cet objectif, et s'accompagner de changement dans l'affect, pas seulement ceux qui construisent les événements mentaux que nous ressentons comme ayant une grande intensité affective, ou que nous appelons des émotions (Barrett, 2016). Alors que les concepts affectifs correspondent à des modèles intéroceptifs, les concepts perceptuels correspondent à des modèles extéroceptifs, et les concepts moteurs correspondent à des modèles proprioceptifs. L'ensemble des concepts générés par le cerveau incarnent différentes échelles de la structure hiérarchique et récursive qui définit son organisation. Ainsi, les concepts d'ordres supérieurs émergent de l'intégration de concepts d'ordres inférieurs, et incarnent donc des modèles prédictifs d'un niveau d'abstraction plus élevé. Par exemple, les concepts qui sous-tendent la capacité d'écrire sont des modèles multimodaux qui intègrent un ensemble de concepts : a) extéroceptifs qui sous-tendent la perception d'objets concrets (ex. papier, stylo, etc.) et abstraits (ex. lettres, mots, etc.); et b) proprioceptifs qui sous-tendent le contrôle moteur permettant la manipulation du stylo afin d'écrire sur la page. C'est en rapport à cette question de

niveau d'abstraction que l'on peut comprendre la différence entre les concepts affectifs et émotionnels.

Les concepts émotionnels sont générés afin de répondre à l'impératif suivant : pour réaliser l'allostasie, le cerveau doit générer des concepts qui lui permettent d'implémenter une modélisation *globale et dynamique* de l'état de l'organisme *en relation à sa situation dans son environnement*. Alors que les concepts affectifs portent uniquement sur les signaux intéroceptifs, et catégorisent l'état homéostatique du corps de l'organisme de manière relativement isolée, les concepts émotionnels portent quant à eux — avec l'aide de l'investissement attentionnel soutenu permis par le potentiel activateur des concepts affectifs — sur un ensemble de signaux à la fois intéroceptifs, proprioceptifs, et extéroceptifs, et modélisent l'état de l'organisme à des échelles spatiotemporelles et informationnelles supérieures (ils sont d'un niveau d'abstraction plus élevé). Cependant, ce qui différencie les émotions des autres concepts multimodaux incarnés par l'activité du cortex n'est pas le fait qu'ils sont associés à une valence affective particulière, puisqu'ils le sont tous⁶, mais plutôt le fait qu'ils sont les modèles prédictifs qui portent directement sur les états affectifs *eux-mêmes*. Dans ce contexte, les concepts émotionnels peuvent être compris comme des métaconcepts multimodaux qui visent la minimisation des erreurs de prédictions intéroceptives véhiculées par les affects en les intégrant à un complexe de prédictions extéroceptives et proprioceptives pertinentes. Selon cette perspective, les émotions incarnent une tentative résoudre *l'incertitude ressentie* incarnée par l'affect (Solms et Friston, 2018), ce qui implique l'utilisation de l'ensemble des processus d'inférences décrits par les approches prédictives.

⁶ Du moins lorsqu'ils sont actifs, puisque l'activité de traitement prédictif qu'ils incarnent n'est rendu possible que par la précision qui leur est accordée par les états affectifs.

À travers l'inférence perceptuelle, principalement réalisée sur la base des entrées sensorielles extéroceptives, le système peut générer un modèle explicatif des états de fait des changements du milieu interne et du milieu externe qui sont à l'origine du changement dans l'affect (ex. « Si j'ai faim, c'est parce que mon corps a besoin que je mange quelque chose. »⁷). Les concepts générés par l'inférence perceptuelle peuvent guider l'inférence active, principalement implémentée par le biais des prédictions proprioceptives, à travers laquelle le système peut mettre en place des actions afin de modifier ses entrées sensorielles afin de se ramener vers son ensemble d'états préférés (ex. « Je peux résoudre cette sensation de faim en allant cuisine manger une banane. »). L'intégration des prédictions extéroceptives et proprioceptives issues de l'inférence perceptuelle et proprioceptive qui peut être utilisée non seulement pour prédire, mais aussi pour prévenir les fluctuations indésirables de l'affect en empêchant les débalancements homéostatiques qui les auraient causés. De plus, les prédictions associées à un concept émotionnel particulier vont pouvoir venir résoudre — c'est-à-dire, expliquer (*explain away*) — certains états affectifs, et ce par le biais des connexions inhibitrices descendant du cortex vers les régions affectives qui vont venir moduler, et moduler la précision accordée aux erreurs de prédiction intéroceptives qu'elles incarnent. Un exemple simple de cet état de fait correspond au délai de gratification qu'un individu peut mettre en place lorsqu'il ressent de la faim, mais que ce n'est pas encore l'heure de manger. Les modèles de haut niveau qui incarnent les plans d'action temporellement étendus peuvent venir dévier l'attention de l'affect homéostatique de faim en l'expliquant, c'est-à-dire en résolvant l'incertitude qui lui est

⁷ Si ces exemples prennent la forme d'une phrase déclarative formulée à la première personne, c'est parce qu'il s'agit de la manière la plus succincte que nous avons trouvé d'explicitier ces processus neurocognitifs autoréférentiels. Cependant, il est important de noter que ces derniers se déroulent en grande partie de manière implicite, et qu'il n'impliquent pas nécessairement une forme de mentation langagière explicite telle que celle suggérée par la formulation de nos exemples. Néanmoins, et tel que nous allons le voir dans le chapitre 3, l'utilisation du langage pourrait jouer un rôle important dans des formes de conceptualisation émotionnelle de haut niveau impliquant un contrôle exécutif supérieur.

associée (ex. « Il ne reste qu'une heure d'ici la pause déjeuner, je vais pouvoir manger bientôt. »).

Comme nous pouvons le constater, la conception dynamique de l'activité neuronale offerte par les approches prédictives remet drastiquement en cause les modèles linéaires classiques des émotions. Selon cette perspective, le cerveau n'est pas un organe passif qui attend les stimulations, il est constamment en train d'anticiper les entrées sensorielles à venir, et n'est jamais vraiment au repos. Au sein des neurosciences affectives, l'adoption de cette compréhension du fonctionnement du cerveau a mené au rejet des modèles classiques qui, sur la base du célèbre schème du « stimulus → réponse », concevaient les émotions comme étant essentiellement des réactions aux événements du monde extérieur. Par opposition, les approches prédictives mettent une emphase radicale sur la dimension « top-down » du traitement de l'information dans le cerveau. Dans ce contexte, les prédictions sont conçues comme des signaux descendants qui sont principalement implémentés par des connexions inhibitrices ayant un pouvoir causal déterminant sur l'activité des échelles inférieures de la hiérarchie prédictive incarnée par le cerveau. Par exemple, ces signaux prédictifs changent littéralement l'activation des neurones sensoriels et moteurs primaires, de manière pouvoir éliciter leur activité, et ce même si aucune entrée sensorielle n'a été reçue par le système (Kok *et al.*, 2016).

En résumé, selon la perspective de Barrett, une émotion correspond à une activité neurocognitive globale impliquant cognition, affect, perception et action afin de générer une modélisation dynamique de l'état de l'organisme par rapport à sa situation dans son environnement, non seulement dans le présent, mais aussi dans le futur. Ce faisant, les émotions sont les événements neurocognitifs qui incarnent les modèles prédictifs les plus intimement liés à la réalisation de l'allostase, puisque cette dernière est définie comme le processus d'autorégulation permettant de maintenir l'homéostasie de l'organisme de manière dynamique, c'est-à-dire, en anticipant et en régulant ses

fluctuations à travers le temps. Ces modèles doivent être flexibles, ils doivent être constamment modifiés pour correspondre aux changements dans le milieu interne et le monde extérieur. C'est ce constat, évident selon nous, qui pousse Barrett à adopter une conception constructiviste des émotions. En effet, il devrait être aisé maintenant de comprendre que si les prototypes d'affects homéostatiques et les patterns d'actions fixes⁸ qui leur sont liés semblent pouvoir être innés, ce n'est pas le cas des émotions en tant que telles, qui doivent être acquises durant le développement de l'organisme et au-delà, à travers les expériences vécues. Terminons cette section avec la citation suivante, qui selon nous résume très bien la position de Barrett :

The brain uses emotion concepts to categorize sensations to construct an instance of emotion. That is, the brain constructs meaning by correctly anticipating (predicting and adjusting to) incoming sensations. Sensations are categorized so that they are (i) actionable in a situated way and therefore (ii) meaningful, based on past experience. When past experiences of emotion (e.g., happiness) are used to categorize the predicted sensory array and guide action, then one experiences or perceives that emotion (happiness). (Barrett, 2016).

L'un des objectifs principaux de la psychothérapie psychodynamique telle que définie par la Neuropsychanalyse consiste en la modification du fonctionnement émotionnel de l'individu. La notion de concept émotionnel développée par Barrett est compatible avec la notion neuropsychanalytique de schéma ou de mode de fonctionnement émotionnel, et nous permet de l'opérationnaliser dans le cadre des approches prédictives du cerveau humain. En effet, Barrett affirme bien que les émotions sont les phénomènes neurocognitifs par le biais desquels l'organisme remplit ses besoins à

⁸ En éthologie, un pattern d'action fixe (*fixed action pattern*) correspond à un schème comportemental instinctif caractérisé par une séquence d'actions indivisibles et stéréotypées dont l'implémentation serait sous-tendue par des substrats neuronaux innés reliés à des stimuli spécifiques (Páez-Rondón *et al.*, 2018).

travers l'interaction avec son environnement, et que ces dernières sont acquises durant le développement, à travers les expériences vécues par l'organisme. Ce faisant, à l'aide de la théorie prédictive des émotions présentée dans cette section, nous pouvons affiner notre compréhension du processus de changement thérapeutique qui sous-tendrait l'efficacité de la psychothérapie psychodynamique : ce dernier consisterait, au moins en partie, en une modification des concepts neurocognitifs qui gouvernent le fonctionnement émotionnel d'un individu. Cependant, dans notre explication, la question restera de savoir comment exactement la psychothérapie psychodynamique peut permettre la modification de ces concepts émotionnels. Afin d'étayer la formulation d'une réponse à cette question, la section suivante portera sur l'investigation des substrats neurologiques qui pourraient sous-tendre la génération et l'implémentation des concepts émotionnels en question.

2.2.4 La neuropsychologie des grands réseaux neuronaux

Une grande partie de notre connaissance sur le cerveau provient du paradigme dit « modulaire », dans lequel les aires cérébrales sont comprises comme des processeurs indépendants réalisant des fonctions cognitives spécifiques. Ce paradigme ne se limite pas au fonctionnement neurologique, mais s'accompagne d'une conception de l'esprit, la psychologie des facultés (Fodor, 1986), comme d'un ensemble de facultés associées à l'accomplissement d'activités mentales diverses (ex. émotion, perception, cognition, action, etc.) Un grand nombre de preuves s'accumulent pour montrer que ce paradigme présente de sérieuses lacunes, et indiquent même qu'il pourrait être erroné. Au sein des neurosciences cognitives, un nouveau paradigme émergent entend dépasser ce localisationnisme en plaçant l'emphase sur la fonction conjointe des aires cérébrales

qui travailleraient ensemble en formant de « grands réseaux » neuronaux (Bressler et Menon, 2010)

C'est l'étude de l'activité spontanée du cerveau en situation de repos, à l'aide de techniques d'imagerie cérébrale telles que l'électroencéphalographie (EEG) ou de l'imagerie fonctionnelle par résonance magnétique (IRMf), qui a permis d'identifier ce que l'on appelle de grands réseaux intrinsèques, qui correspondent à l'organisation spontanée de l'activité cérébrale indépendante du traitement de tâches cognitives spécifiques. La plupart de ces grands réseaux intrinsèques sont aussi actifs durant des opérations cognitives spécifiques, ce qui suggère que le cerveau humain pourrait être intrinsèquement organisé en différents réseaux fonctionnels. Aujourd'hui, les grands réseaux fonctionnels sont conçus comme des formes d'organisation de l'activité du cerveau humain qui émergent de l'activation simultanée et coordonnée d'un ensemble de régions corticales et sous-corticales cérébrales spécifiques, et qui sous-tendraient l'accomplissement d'une grande variété de fonctions cognitives.

La littérature en neuropsychologie semble suggérer de manière relativement unanime que l'accomplissement de fonctions et d'activités psychologiques complexes, telles que celles qui sous-tendent les émotions et leur régulation, repose sur l'activation globale du cerveau correspondant à l'échelle neurologique des grands réseaux neuronaux (Lindquist et Barrett, 2012). Dans le même ordre d'idée, nous pensons que les grands réseaux neuronaux constituent le niveau d'organisation le plus pertinent pour l'explicitation de l'espace d'échelle des phénomènes en jeu dans la psychothérapie psychodynamique. Dans ce qui suit, nous allons procéder à une brève revue des grands réseaux qui ont été découverts par la recherche en neuropsychologie, et qui sont pertinents pour l'explication de la psychothérapie psychodynamique.

Le système limbique : principalement composé du lobe temporal médian (MTL ; incluant l'hippocampe et le cortex enthorinal), du cortex cingulaire antérieur (ACC),

du cortex orbitofrontal, de l'amygdale, du striatum ventral et de la substance grise périaqueducale, *le système limbique* est le réseau neuronal typiquement associé aux émotions (Thomas Yeo *et al.*, 2011). Son activité est associée de manière consistante avec l'expérience affective et émotionnelle (Kober *et al.*, 2008 ; Lindquist *et al.*, 2012), la mémoire autobiographique (Spreng *et al.*, 2008), l'expérience de la douleur (Lamm *et al.*, 2011) et la mise en place de comportements motivés (Grillner *et al.*, 2005). Selon Lindquist & Barrett (2012) la fonction du système limbique serait de représenter et de réguler les états viscéromoteurs de l'organisme afin de générer la tonalité affective basique (plaisir, déplaisir, et excitation) qui caractérise toute expérience consciente et dirige les comportements basiques d'exploration ou de retrait. Comme nous pouvons le remarquer, cette fonction est très similaire à celle que Friston et Solms & Friston (2018) attribuent à l'ARAS, ce qui n'est pas étonnant puisque le système limbique est l'une des structures cérébrales principales à travers lesquels l'ARAS exerce son influence sur le cortex (Brudzynski, 2013 ; Siposan et Aliu, 2014). Bien que le système limbique constitue bien un réseau intrinsèque, dans le sens où il présente une connectivité robuste indépendante à la réalisation de tâches cognitives spécifiques, son extension relativement limitée fait que l'on ne peut pas le considérer on ne peut pas le considérer comme un « grand » réseau au même titre que les réseaux que nous allons présenter par la suite. Néanmoins, il était selon nous nécessaire de le présenter ici à cause de son rôle critique dans les émotions.

Le réseau de contrôle exécutif : principalement composé du cortex préfrontal dorsolatéral (dlPFC), du cortex pariétal postérieur (PPC), du précuneus, du sulcus pariétal inférieur, et du cortex cingulaire médian, le réseau de contrôle exécutif (aussi appelé *central executive network* ; CEN) serait le grand réseau responsable des fonctions exécutives (Thomas Yeo *et al.*, 2011). L'activité du CEN est associée de manière consistante avec l'engagement dans des tâches cognitives impliquant un investissement soutenu de l'attention vers des stimuli extéroceptifs, à l'alternance entre ces tâches, et à l'état de vigilance à des indices perceptuels indiquant l'arrivée de

stimuli externes (Crone *et al.*, 2006 ; Fan *et al.*, 2005). Le CEN jouerait un rôle crucial dans la capacité à maintenir et à manipuler l'information dans la mémoire de travail, dans la résolution de problème basée sur des règles, ainsi que dans la prise de décision dans le contexte de comportements orientés par des buts (Crone *et al.*, 2006 ; Fan *et al.*, 2005).

Le réseau du mode par défaut : principalement composé du cortex préfrontal médian (mPFC), le cortex cingulaire postérieur (PCC), le lobule pariétal inférieur (IPL), les lobes temporaux médians (MTL ; qui incluent l'hippocampe et le cortex entorhinal), le gyrus angulaire, et sulcus temporal supérieur, le réseau du mode par défaut (*Default Mode Network*; DMN) est le premier des grands réseaux intrinsèques à avoir été découvert à l'aide de l'IRMf (Greicius *et al.*, 2003). S'il a été appelé ainsi, c'est parce que lors de sa découverte, il a été associé à un état de repos non contraint, c'est-à-dire au mode de cognition dans lequel nous nous engageons le plus souvent lorsque nous ne sommes pas occupés par la réalisation d'une tâche spécifique : une activité mentale majoritairement autoréférentielle, relativement indépendante des stimuli externes (*mind wandering*), et principalement composée du rappel d'expériences passées et d'une projection dans le futur (Raichle, 2015 ; Raichle *et al.*, 2001). En contraste avec le CEN, le DMN est typiquement désactivé durant la plupart des tâches cognitives basées sur des stimuli externes (Raichle *et al.*, 2001), et la recherche a montré que ces deux réseaux ont la particularité d'être fonctionnellement anti-corrélés, ce qui veut dire que l'activation de l'un s'accompagne généralement par la désactivation de l'autre (et vice versa).

Depuis sa découverte, les différents nœuds du DMN ont été associés à divers processus cognitifs qui sous-tendraient non seulement l'activité mentale qui constitue notre « mode par défaut », mais aussi la réalisation de tâches cognitives spécifiques. Les nœuds du PCC, de l'hippocampe, et du gyrus angulaire ont été typiquement associés avec le rappel au niveau de la mémoire épisodique (Sestieri *et al.*, 2011 ; Vannini *et al.*,

2011), la mémoire autobiographique (Dastjerdi *et al.*, 2011 ; Spreng *et al.*, 2008), et la mémoire sémantique reliée à la pensée interne (Binder *et al.*, 2009), alors que les nœuds du mPFC ont été associés aux processus cognitifs reliés au Soi et à la cognition sociale (Amodio et Frith, 2006 ; Spreng *et al.*, 2008), la prise de décision basée sur les valeurs personnelles (Rangel *et al.*, 2008), et la régulation émotionnelle (Etkin *et al.*, 2011). Ainsi, pris dans son ensemble, le DMN consisterait en un système intégré dont l'activité serait responsable de différents aspects des processus mentaux autoréférentiels (Menon, 2011) tels que l'autobiographie, d'autorégulation émotionnelle, le Soi et la cognition sociale. Ces processus mentaux correspondent directement à ceux que nous avons associés à la pratique de la psychothérapie psychodynamique dans notre explanandum. Cela fait du DMN la structure cérébrale la plus prometteuse pour l'explication processus neurocognitifs qui sous-tendraient les phénomènes en jeu dans la psychothérapie psychodynamique. Pour cette raison l'espace d'échelle de l'organisation neurocognitive humaine incarné par le DMN constituera le cœur du modèle explicatif que nous allons tenter d'élaborer au chapitre 3 de ce mémoire.

Ce qui est particulièrement intéressant pour nous est que le DMN a été systématiquement associé à la théorie de l'esprit et à la capacité de mentalisation (Denny *et al.*, 2012 ; Frith *et al.*, 2003 ; Lombardo *et al.*, 2009 ; Spreng *et al.*, 2008 ; Spunt *et al.*, 2010, 2015), de telle manière que Barrett et Satpute (2013) ont choisi de l'appeler le « réseau de la mentalisation » (*mentalizing network*). En continuité avec la théorie prédictive des émotions développée par Barrett, la fonction générale qu'ils attribuent à ce réseau est une activité de *conceptualisation*. Cette dernière consisterait en un processus de génération de concepts neurocognitifs, des modèles prédictifs qui implémentent une simulation des expériences passées (cf. souvenirs épisodiques ou connaissances catégoriques) pour construire un modèle mental du passé, du présent, et du futur (Barrett et Satpute, 2013 ; Lindquist et Barrett, 2012). Dans le cadre de la mémoire autobiographique ou de la cognition abstraite (conceptuelle), ce processus

permettrait d'implémenter une simulation des expériences sensorimotrices passées. Dans le cadre de la perception, ce processus participerait à intégrer les connaissances préalables afin d'attribuer un sens aux objets perceptuels. Dans le cadre des émotions, ce processus aiderait à intégrer les sensations intéroceptives avec le contexte extéroceptif afin d'élaborer un modèle dynamique de l'état homéostatique de l'organisme (allostasie) par rapport à sa situation environnementale (Lindquist et Barrett, 2012).

Autrement dit, le DMN serait incarnerait l'échelle d'organisation cérébrale globale dont l'activité serait responsable de la génération et de l'implémentation des concepts émotionnels. À la section précédente, nous avons formulé l'hypothèse selon laquelle le changement du fonctionnement psychologique d'un individu qui s'opère durant la psychothérapie psychodynamique reposerait sur un mécanisme sous-jacent de modification des concepts neurocognitifs qui sous-tendent son fonctionnement émotionnel. Si l'on associe cette hypothèse à la thèse précédente sur la fonction du DMN, alors il s'ensuit que le changement thérapeutique élicité par la psychothérapie psychodynamique devrait reposer sur une modification de l'organisation du DMN.

Le réseau de la saillance : principalement composé du cortex cingulaire antérieur (ACC), de l'insula dorsal antérieur bilatéral, de l'opercule frontal, et fortement connecté à des aires sous corticales impliquant l'amygdale, la substance noire et le thalamus, le réseau de la saillance (*Saliency Network*; SN) est intimement lié au système limbique et à l'ARAS (Seeley *et al.*, 2007). L'activation du SN est associée de manière consistante avec l'expérience affective et émotionnelle (Kober *et al.*, 2008 ; Lindquist *et al.*, 2012), l'aversion (Hayes et Northoff, 2011), la pensée envers un partenaire romantique (Bartels et Zeki, 2000), la douleur (Lamm *et al.*, 2011), l'orgasme féminin (Bianchi-Demicheli et Ortigue, 2007), ainsi que l'attention et le langage (Nelson *et al.*, 2010). Selon Menon (2011), la recherche récente (Lovero *et al.*, 2009 ; Menon et Uddin, 2010 ; Seeley *et al.*, 2007 ; Sridharan *et al.*, 2008) suggèrerait

que la fonction de ce réseau serait d'identifier, au sein de l'ensemble des stimuli intra et extrapersonnels, ceux qui sont le plus pertinents pour la régulation de l'homéostasie, et de diriger l'attention sur ces derniers (les rendant saillants) afin de guider le comportement et la cognition. Sur la base d'une revue extensive de la littérature, Barrett et Satpute (2013) affirment que le SN serait composé de deux sous-réseaux : a) un sous-réseau ventral qui conceptualiserait l'information intéroceptive sous la forme d'états affectifs discrets (plus spécifiques que la simple tonalité affective globale générée par le système limbique) ; et b) un sous réseau dorsal qui utiliserait ces états affectifs pour guider l'attention et le comportement.

Si nous avons choisi de présenter le SN en dernier, c'est à cause du rôle central qu'il jouerait dans l'activation des autres grands réseaux à travers la direction des ressources attentionnelles. En effet, la recherche sur l'influence des nœuds du SN sur d'autres régions cérébrales a montré que le cortex insulaire antérieur joue un rôle causal dans l'activation alternative du CEN ou du DMN (Menon et Uddin, 2010 ; Sridharan *et al.*, 2008 ; Zhou *et al.*, 2018). En effet, le CEN et le DMN seraient en compétition à travers les types de tâches cognitives et les différentes modalités sensorielles, et sont considérés comme dirigeant l'attention vers les mondes externe et interne, respectivement (Menon, 2011). Le rôle causal attribué au SN dans la direction de l'activité globale du cerveau à travers l'allocation des ressources attentionnelles en fonction des exigences du processus de régulation homéostatique de l'organisme est en directe continuité avec la thèse de Solms et Friston (2018) sur la conscience affective présentée précédemment. Dans le cadre des approches prédictives, l'activité du SN peut être comprise comme l'une des voies privilégiées de l'implémentation du processus d'allocation de la précision aux sensations et au concepts (abstractions) neurocognitifs les plus pertinents pour la réalisation de l'allostasie. Carhart-Harris et Friston (2010) affirment que dans le cadre des approches prédictives, les grands réseaux intrinsèques correspondent aux niveaux supérieurs de la hiérarchie prédictive incarnée par le cerveau, qui suppriment l'énergie libre des niveaux inférieurs, en

minimisant les erreurs de prédictions qu'ils véhiculent à l'aide de prédictions *top-down* (implémentées par des connexions descendantes inhibitrices) qui les expliquent (*explain away*).

La conception globale de l'activité cérébrale qui se dessine à travers l'intégration du FEP et des approches prédictives avec la neuropsychologie des grands réseaux neuronaux est la suivante : le cerveau humain est un organe engagé dans une activité de constante de minimisation des erreurs de prédiction reliées aux entrées sensorielles qu'il reçoit du milieu interne et externe à l'organisme dont il fait partie, tout ceci dans le but de minimiser l'énergie libre des états internes de ce dernier en le maintenant dans les normes homéostatiques définies par son phénotype. La toile de fond de cette activité correspondrait à un mode par défaut caractérisé par un processus de conceptualisation à travers lequel les échelles supérieures de la hiérarchie cérébrale implémentent des modèles prédictifs générés à partir des expériences passées afin d'expliquer et de prédire les affects, qui à travers une synthèse des erreurs de prédictions intéroceptives, encodent l'incertitude ressentie par l'organisme vis-à-vis de l'intégrité de ses états internes. Cependant, un organisme tel qu'un être humain vit dans un environnement complexe et imprévisible, et il est constamment bombardé d'entrées sensorielles extéroceptives ayant une pertinence pour le processus de réalisation de son homéostasie. Lorsque les erreurs de prédictions suscitées par ces dernières mènent l'incertitude vis-à-vis des états internes à dépasser un certain seuil, et qu'il n'est plus prédit que l'activité de conceptualisation du mode par défaut à elle seule puisse minimiser l'énergie libre du système, le cerveau opère une transition vers une activité d'inférence active à travers laquelle il va coordonner l'interaction de l'organisme avec le monde extérieur afin de rétablir les conditions environnementales propices au maintien de son organisation.

2.3 L'échelle écologique et socioculturelle de la cognition humaine

L'une des hypothèses centrales de la psychologie psychodynamique consiste à affirmer que le fonctionnement psychologique typiquement humain est issu d'un processus de conditionnement instauré par l'environnement social de l'individu durant son développement et au-delà. Nous terminerons ce chapitre par l'introduction de deux modèles théoriques issus de la philosophie de l'esprit contemporaine qui nous offrent des outils conceptuels qui nous permettront d'explicitier la dimension écologique et socioculturelle de la cognition humaine qui sous-tend le fonctionnement de la psychothérapie psychodynamique.

2.3.1 Le « Skillfull Intentionality Framework »

Afin de compléter notre modèle de la cognition humaine, nous avons choisi de nous appuyer sur une approche contemporaine, le *Skillful Intentionality Framework* (SIF ; Rietveld *et al.*, 2018) issue de l'intégration du FEP avec la philosophie de l'esprit énonciviste et la psychologie écologique, des approches qui possèdent toutes deux des appareils conceptuels très pertinents pour notre projet.

L'approche énoncive de la cognition est originellement issue de l'application de la conception autopoïétique du vivant à l'étude de la cognition (Bruin *et al.*, 2018), une tentative visant à considérer la cognition comme une extension du processus d'autoproduction (autopoïèse) qui définit les systèmes vivants. Elle fut développée de concert avec la cognition incarnée (*embodied cognition*; Varela *et al.*, 2000), une

position selon laquelle les processus cognitifs dépendent de la structure morphologique et anatomique du corps de l'organisme qui les performe. L'aspect principal de l'approche énaïve de la cognition est l'emphase pragmatiste sur l'action adaptative : l'ensemble des significations qui composent le vécu d'un agent cognitif émergent d'un processus d'interaction continu entre un organisme et son environnement.

La psychologie écologique (Gibson, 2011) est une approche qui découle de l'idée que la compréhension de l'activité cognitive (en particulier la perception) d'un organisme doit nécessairement passer par une analyse approfondie de la niche écologique dans laquelle il vit, et qui correspond à la forme de vie particulière qu'il incarne (Rietveld *et al.*, 2018). Cette discipline s'organise autour du concept d'affordances, qui correspondent aux possibilités d'actions issues de l'interaction entre les habiletés d'un agent cognitif et les attributs pertinents de son environnement matériel.

Le SIF est une approche philosophique intégrative qui vise à fournir un nouveau cadre conceptuel pour rendre justice au phénomène complexe de la cognition incarnée. Ce cadre conceptuel est principalement caractérisé par la combinaison d'une version enrichie de la notion d'affordance (psychologie écologique) avec une emphase sur l'action adaptative (énaïvisme). L'intentionnalité habile (*skillful intentionality*) y est définie comme l'engagement sélectif et simultané avec de multiples affordances dans le contexte d'une situation concrète (Rietveld *et al.*, 2018). À travers le concept de paysage d'affordances, le SIF capture la manière dont les possibilités d'actions se présentent à un organisme comme un échafaudage d'affordances interdépendantes et imbriquées les unes dans les autres (Bruineberg et Rietveld, 2014).

Les partisans du SIF considèrent que l'échelle appropriée à l'étude de la cognition est celle du système cerveau-corps-environnement, et voient dans le cadre métathéorique du FEP des ressources prometteuses pour rendre compte des dynamiques d'autorégulation de ce système. Selon eux, il est possible de comprendre le processus

cérébral de minimisation de l'erreur de prédiction, extension de l'évitement de la surprise, comme un mécanisme ayant pour but l'implémentation d'un couplage structurel robuste entre l'organisme et l'environnement (Bruineberg et Rietveld, 2014). Ils précisent leur position en affirmant que l'organisme n'incarne pas un modèle de la structure causale de l'environnement en soi, mais plutôt un modèle de la structure dynamique du champ d'affordances, c'est-à-dire des potentialités d'action (ou d'inférence active), des actions qui doivent être comprises comme des extensions du processus de réalisation de l'allostase. Le concept d'affordance tel que compris par le SIF nous sera utile dans notre tentative de description de la manière dont les concepts émotionnels, ainsi que la capacité de mentalisation, sont transmis durant le développement (ou la psychothérapie) à travers les interactions qui ont lieu entre un individu et son environnement social proximal.

2.3.2 Le modèle des affordances culturelles

Dans un article récent, Ramstead *et al.* (2016) tentent de compléter le SIF en adressant la question des affordances particulières qui émergent au sein des niches socioculturelles complexes spécifiques à l'espèce humaine. En intégrant le SIF aux avancées récentes en anthropologie cognitive et évolutionnaire qui portent sur la coévolution de la culture et des gènes (Henrich, 2016) ainsi que sur nature collaborative de l'espèce humaine (Hrdy, 2009 ; Tomasello, 2014), ils ont entrepris de développer une conception des affordances culturelles et des mécanismes qui les sous-tendent. Leur travail s'inscrit dans le contexte de la recherche pour « les origines naturelles du contenu » (Hutto et Satne, 2015) qui vise à fournir une explication naturaliste de l'émergence du contenu sémantique, c'est-à-dire, de l'évolution (dans la phylogénie) et l'acquisition (dans l'ontogénie) de contenu représentationnel et propositionnel.

Conformément aux principes de la psychologie écologique, la culture est considérée ici comme une échelle environnementale spécifique qui est définie par un domaine informationnel qui lui est propre. Cet environnement est essentiellement constitué des pratiques sociales structurées et des stratégies de construction de niche qui sont propres à l'espèce humaine, et qui encodent ou « contiennent » l'information culturelle. Les affordances culturelles correspondent à la possibilité pour l'organisme de percevoir et d'actualiser les opportunités d'action offertes par cette échelle environnementale spécifique.

L'idée est que les humains deviennent sensibles aux affordances culturelles de leur niche socioculturelle à travers la modulation de l'attention qui est rendue possible par la participation immersive dans certaines formes de pratiques culturelles structurées, et qui sont appelées des « régimes d'attention partagée ». Un régime d'attention est une pratique collective qui signale la saillance (ou l'inverse) de certaines formes de stimuli et de situations. Selon le FEP, l'attention est un mécanisme cérébral central qui a pour fonction de moduler l'importance (la pondération) accordée aux erreurs de prédiction et aux prédictions du système prédictive dans son activité de modélisation prédictive subséquente. Ce faisant, la direction de l'attention détermine la dynamique globale du système cognitif et guide le déploiement des boucles d'action-perception (inférence active et perceptuelle) mises en place un agent cognitif. En dirigeant les processus attentionnels, les régimes d'attention mènent les agents à intérioriser les modèles prédictifs qui leur permettent de s'intégrer à la culture du groupe socioculturel auquel ils appartiennent. En couvrant l'échelle socioculturelle de la cognition humaine, la dernière nécessaire à l'explication de la psychothérapie psychodynamique, cette section conclut le deuxième chapitre de notre mémoire de maîtrise.

2.4 Synthèse

Dans ce chapitre, nous avons présenté un ensemble de modèles théoriques qui décrivent les multiples échelles de l'organisation des systèmes neurocognitifs humains à partir d'un principe organisationnel commun à l'ensemble de ces systèmes physiques particuliers que sont les systèmes vivants : l'impératif de réduction de leur entropie interne.

Tout d'abord, la théorie de l'autopoïèse nous a offert une compréhension initiale des êtres vivants comme des systèmes physiques définis par leur capacité à générer et à maintenir les processus qui définissent leur propre organisation à travers le temps tout en affirmant leur identité en installant une distinction claire avec la niche écologique dont ils émergent (Maturana et Varela, 1980). Selon cette théorie, l'histoire des interactions récurrentes entre un système vivant et le système écologique dont il dépend pour s'approprier les ressources nécessaires au maintien des processus qui le définissent, mène à l'instauration d'un couplage structurel à travers lequel ces deux systèmes en viennent à développer une congruence structurelle, c'est-à-dire qu'ils façonnent mutuellement les processus qui définissent leurs organisations respectives (Bruin et al., 2018). Ensuite, le principe de l'énergie libre (FEP) nous a montré la manière dont les systèmes vivants accomplissent l'exploit de maintenir leur organisation à travers le temps : ils réduisent activement la dispersion ou le désordre de leurs états sensoriels et physiologiques, et réalisent cela en minimisant leur *énergie libre variationnelle*, une quantité informationnelle qui encode la limite supérieure de l'entropie de leurs échanges sensoriels avec leur environnement (Friston, 2010). Le FEP affirme que pour qu'un système résiste à l'érosion entropique et se maintienne dans un ensemble d'états limité (pour posséder une homéostasie généralisée) il doit instancier un modèle statistique et causal de sa relation avec son éconiche (Ramstead

et al., 2017). En d'autres termes, un organisme ne fait pas juste encoder un modèle du monde, il est un modèle du monde — une transcription physique des régularités causales dans son éconiche qui a été sculptée par des interactions réciproques (couplage structurel) entre son auto-organisation et la sélection naturelle (pressions environnementales) à travers le temps.

Le modèle ontologique de la neuroéthologie variationnelle nous a permis de nous représenter les processus qui définissent les différentes échelles spatiotemporelles de l'auto-organisation des systèmes vivants comme différents moyens mécaniquement hétérogènes de minimiser l'énergie libre de leurs états internes (Ramstead *et al.*, 2017). Les approches prédictives nous ont fourni une description des mécanismes à travers lesquels le cerveau humain réalise le processus de minimisation de l'énergie libre : il incarne une machine à inférence qui cherche à prédire les entrées sensorielles provenant du milieu interne (intéroception) et externe (extéroception) à l'organisme auquel il appartient en générant des modèles probabilistes de leurs causes distales (Hohwy, 2013). Le cerveau implémente l'impératif de minimisation de l'énergie libre en cherchant à minimiser ses erreurs de prédictions, qui correspondent à la différence entre les prédictions qu'il génère et les entrées sensorielles effectives qu'il reçoit. Ici, la fonction du cerveau est définie comme la réalisation de l'allostasie de l'organisme, qui correspond au processus dynamique de maintien des états internes de ce dernier dans les normes homéostatiques définies par son phénotype (Barrett, 2016 ; Sterling, 2012).

Solms et Friston (2018) nous ont proposé une conception de la *conscience*, c'est-à-dire, de *l'expérience subjective* qui accompagne la cognition humaine, comme étant intrinsèquement liée à ce que nous appelons les *affects*, qui correspondraient à la manière particulière à travers laquelle le système nerveux central des animaux vertébrés encode les erreurs de prédictions *intéroceptives* qui l'informent des variations d'états homéostatiques de leur corps. Selon ce modèle, les affects correspondent à des signaux qui encodent *l'incertitude ressentie* par l'organisme vis-à-vis du processus de

réalisation de son l'allostasie, une incertitude qui va constituer le principal moteur de l'activité du système neurocognitif dans son ensemble. D'après Barrett (2016), la manière principale dont le cerveau humain optimise le processus de réalisation l'allostasie de l'organisme auquel il appartient consiste à générer et à implémenter des *concepts émotionnels* : des modèles prédictifs multisensoriels qui expliquent les erreurs de prédiction intéroceptives, extéroceptives et proprioceptive afin générer une simulation dynamique permettant de prédire et de réguler l'état de l'organisme par rapport à sa situation dans son environnement, non seulement dans le présent, mais aussi dans le futur. Grâce au paradigme émergeant de la neuropsychologie des grands réseaux, nous avons pu comprendre que la génération et l'implémentation des concepts émotionnels, de même que les autres formes de cognition de haut niveau qui sont impliquées dans la psychothérapie psychodynamique (ex. mentalisation), sont fort probablement sous-tendue par l'activité les échelles supérieures de la hiérarchie cérébrale, qui supervisent et coordonnent l'activité de l'ensemble du cerveau.

Le « Skillfull intentionality framework » (Rietveld *et al.*, 2018) nous a permis de réaliser que l'échelle appropriée à l'étude des formes de cognition impliquées dans la psychothérapie psychodynamique est probablement celle du système cerveau-corps-environnement. Il nous a permis de comprendre le processus cérébral de minimisation de l'erreur de prédiction comme un mécanisme ayant pour but l'implémentation d'un couplage structurel robuste entre l'organisme et son environnement. Afin d'instaurer ce couplage structurel, l'organisme n'incarnerait pas un modèle de la structure causale de l'environnement en soi, mais plutôt un modèle de la structure dynamique d'un paysage *d'affordances*, des possibilités d'actions issues de l'interaction entre les habiletés d'un agent cognitif et les attributs pertinents de son environnement matériel, des actions ayant pour but de réaliser l'allostasie de l'organisme. Enfin, le modèle des affordances culturelles (Ramstead *et al.* 2016) nous ont offert une compréhension des mécanismes à travers lesquels les systèmes neurocognitifs humains acquièrent la capacité de s'engager dans le paysage d'affordances leur permettant d'instaurer un

couplage structurel avec l'échelle socioculturelle de leur niche écologique. Ce mécanisme correspondrait à l'engagement dans des régimes d'attention partagée, des pratiques collectives qui signalent la saillance (ou l'inverse) de certaines formes de stimuli et de situations. Ainsi, le modèle des affordances culturelles nous offre un mécanisme, l'installation de régimes d'attention partagés, à l'aide duquel nous pouvons expliquer comment l'environnement socioculturel peut influencer l'organisation du système neurocognitif d'un individu. Ce mécanisme constituera l'une des charnières centrales de notre explication de la manière dont les interactions interpersonnelles qui ont lieu dans le cadre de la psychothérapie psychodynamique peuvent mener à la modification du fonctionnement émotionnel d'un individu.

CHAPITRE III

EXPLICATION : UN NOUVEAU CADRE THÉORIQUE POUR LA PSYCHOTHÉRAPIE PSYCHODYNAMIQUE ISSU DES NEUROSCIENCE COGNITIVES CONTEMPORAINES

Dans ce troisième et dernier chapitre de notre mémoire, nous allons intégrer des travaux récents réalisés dans le cadre des approches théoriques présentées dans notre *explanans*, et qui pris ensemble vont nous permettre d'appuyer, de modifier, ou d'étendre les hypothèses de la Neuropsychanalyse et de la théorie de la mentalisation, afin d'en arriver à la meilleure explication de la psychothérapie psychodynamique qui puisse être formulée dans le langage des sciences de l'esprit contemporaines. Ce faisant, une grande partie de ce chapitre consistera en l'exposé de modèles théoriques qui portent de manière plus directe sur les phénomènes en jeu dans la psychothérapie psychodynamique. Ces derniers peuvent être considérés comme constituant une sorte *d'explanans intermédiaire*, faisant le pont entre le notre *explanans* et notre *explanandum*. C'est seulement une fois ces modèles présentés que nous pourrons procéder à leur synthèse afin de proposer une reconceptualisation de la pratique de la psychothérapie psychodynamique.

3.1 L'espace d'échelle de la psychothérapie psychodynamique

3.1.1 L'Ego, le réseau du mode par défaut, et le Soi psychologique

Dans un article intitulé « The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas », Carhart-Harris et Friston (2010) allient le paradigme du FEP à la neuropsychologie des grands réseaux neuronaux pour avancer une défense de la validité de construit des notions psychanalytiques de processus primaires et secondaires. Plus spécifiquement, ils proposent que les descriptions que Freud faisait des processus primaires et secondaires soient consistantes avec l'activité auto-organisée dans le système hiérarchique cortical, et que sa conception de la structure psychique de l'Ego est consistante avec les fonctions du réseau du mode par défaut (DMN) et ses échanges réciproques avec les systèmes cérébraux subordonnés. Dans les paragraphes qui suivent, nous allons résumer les points centraux de la proposition de Carhart-Harris et Friston (2010), qui va constituer l'hypothèse centrale à travers laquelle nous allons tenter de définir l'espace d'échelle correspondant aux phénomènes en jeu dans la psychothérapie psychodynamique.

Comme nous l'avons vu au chapitre I, Freud faisait l'hypothèse que l'évolution ontogénétique et phylogénétique d'une cognition humaine mature (adulte) et saine dépendrait de la formation d'un équilibre entre deux échelles d'organisation de l'appareil psychique, le Ça et l'Ego, implémentant deux modes de cognition antagonistes : 1) les processus primaires, organisés autour de motivations instinctuelles (pulsions) visant l'accomplissement immédiat des besoins innés de l'organisme ; et 2) les processus secondaires, visant l'inhibition et la régulation des processus primaires afin qu'ils se traduisent en une cognition et un comportement fonctionnels et adaptatifs. Carhart-Harris et Friston (2010) affirment que cette description que Freud fait de

l'organisation de l'appareil psychique est remarquablement consistante avec les modèles contemporains de la cognition basés sur les approches prédictives du cerveau humain, dans lesquels les connexions descendantes (*top-down*) des aires corticales supérieures travaillent pour minimiser l'énergie libre des aires cérébrales inférieures.

De manière importante, Carhart-Harris et Friston (2010) affirment que le principe de l'énergie libre, tel qu'appliqué au fonctionnement du cerveau à travers le paradigme des approches prédictives, constitue une récapitulation du — ce qui veut dire qu'il est formellement équivalent au — principe de constance formulé par Freud dans son projet pour une psychologie scientifique (Freud, 1954). Plus précisément, ils affirment que le processus (décrit par Freud) de réduction de la somme d'excitation contenue dans l'appareil psychique *est exactement la même chose*⁹ que le processus de minimisation des erreurs de prédictions du système nerveux (décrit par les approches prédictives). Ils expliquent cette correspondance par le fait que le principe de constance et le principe de l'énergie libre sont tous les deux dérivés des hypothèses de la psychophysique du 19^e siècle, et en particulier des travaux de Hermann von Helmholtz et de Gustav Fechner. Leurs affirmations reposent sur une revue¹⁰ extensive et méticuleuse de l'œuvre de Freud, et sur la mise en relation systématique de ses thèses (corroborées par des citations) avec celles du FEP et des approches prédictives, dont il est important de rappeler que Karl Friston constitue le principal architecte. Voyons maintenant la

⁹ « ... the process of minimizing 'the sums of excitation' is exactly the same as minimizing the sum of squared prediction-error or free-energy in Helmholtzian schemes.» (p.1270 ; Carhart-Harris et Friston, 2010)

¹⁰ Cette revue est fournie en tant matériel supplémentaire à leur article intitulé « *The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas* » (Carhart-Harris et Friston, 2010), et cité précédemment.

manière exacte dont Carhart-Harris et Friston identifient les fonctions de l'Ego à celles du réseau du mode par défaut.

Dans leur description de la manière dont le DMN pourrait implémenter les processus secondaires, Carhart-Harris et Friston se concentrent sur la suppression de l'activité limbique et paralimbique par le cortex préfrontal médian (mPFC ; nœud frontal du DMN), qu'ils associent à la suppression de l'activité endogène par l'Ego. En effet, ils expliquent qu'il existe un large champ de la littérature qui démontre cette fonction inhibitrice du mPFC sur les aires limbiques et paralimbiques (Etkin *et al.*, 2011 ; Hariri *et al.*, 2000 ; Milad *et al.*, 2006 ; Milad et Quirk, 2002 ; Phelps *et al.*, 2004 ; Phillips *et al.*, 2003 ; Rosenkranz et Grace, 2002). Ils ajoutent que, de manière consistante avec la description que Freud fait des processus secondaires, cette inhibition a été systématiquement associée à la régulation des émotions (Beauregard *et al.*, 2001 ; Bremner *et al.*, 1999 ; Dougherty *et al.*, 2004 ; Hopper *et al.*, 2007 ; Lanius *et al.*, 2002 ; Pietrini *et al.*, 2000 ; Reinders *et al.*, 2003, 2006 ; Shin, 2006 ; Shin *et al.*, 2004) et au contrôle de l'impulsivité (Anderson *et al.*, 1999 ; Davidson *et al.*, 2000 ; Grafman *et al.*, 1996 ; Kaplan-Solms et Solms, 2002 ; Pietrini *et al.*, 2000 ; Solms et Turnbull, 2018).

Dans le même ordre d'idée, Carhart-Harris et Friston (2010) associent les processus primaires à une activité limbique et paralimbique débridée qui, étant privée de la modulation top-down efficace du DMN, peut influencer l'activité des niveaux intermédiaires de la hiérarchie, et ainsi entraîner une altération des processus cognitifs comme la perception ou la pensée. En continuité avec Freud, ils considèrent que les états mentaux non ordinaires sont la meilleure porte d'entrée pour observer les particularités cognitives et neurophysiologiques des processus primaires de manière isolée. Ces derniers comprennent : le sommeil REM, les états psychotiques aigus, l'épilepsie de l'aura du lobe temporal, et les états hallucinogènes provoqués par les drogues. Ils partagent tous une phénoménologie cohérente qui comporte :

l'atemporalité ; des sensations de peur et de crainte irrationnelles ; des distorsions perceptuelles/hallucinations visuelles ; des phénomènes de déjà vu ou de rappel vif ; et un dérangement du sens de soi.

Freud affirmait que l'instance psychique de l'Ego avait pour fonction de réduire l'ensemble de l'excitation, à la fois endogène et exogène, contenue dans l'appareil psychique. Cependant, il est important de noter que Freud identifiait l'excitation endogène à l'intéroception, et l'excitation exogène à l'extéroception. Or, selon Carhart-Harris et Friston, en termes neurobiologiques, les signaux exogènes correspondent aux signaux intéroceptifs et extéroceptifs venant des aires sensorielles thalamiques et unimodales, qui convient les signaux sensoriels aux structures des lobes polymodaux et temporaux médians. Quant à eux, les signaux endogènes peuvent être considérés comme équivalents aux erreurs de prédictions bottom-up subséquentes (excitation) émergeant dans les aires sensorielles primaires et dans les systèmes limbique et paralimbique, qui sont transmises aux aires corticales polymodales supérieures qui comprennent les nœuds qui composent les grands réseaux comme le réseau du contrôle exécutif (CEN), le réseau de la saillance (SN) et le réseau du mode par défaut (DMN). En effet, les grands réseaux sont tous composés d'aires corticales associatives de haut niveau qui sont éloignées (spatialement) des aires sensorielles primaires, et qui encodent les représentations (cf. concepts) du plus haut niveau d'abstraction. C'est ce qui pousse Carhart-Harris et Friston (2010), tel que nous l'avons mentionné au chapitre précédent, à affirmer que les grands réseaux incarnent l'échelle supérieure de la hiérarchie prédictive incarnée par le cerveau, et qu'ils implémentent les modèles prédictifs les plus élevés.

D'après Carhart-Harris et Friston (2010), le DMN se situe au niveau le plus élevé de la hiérarchie prédictive incarnée par le cerveau, et sa fonction est de minimiser l'énergie libre des échelles intermédiaires, incluant les autres grands réseaux, à travers la génération de prédictions descendantes qui expliquent (suppriment) les erreurs de

prédictions ascendantes qu'elles lui transmettent. Selon eux, l'Ego est la propriété qui émerge des processus qui définissent cette échelle supérieure de l'auto-organisation des systèmes neurocognitifs humains. Afin d'appuyer cette proposition, Carhart-Harris et Friston (dans Fotopoulou *et al.*, 2012) procèdent à l'énumération suivante des caractéristiques uniques du réseau du mode par défaut : (1) il a plus de connexions avec les autres aires cérébrales que n'importe quel autre réseau neuronal, ce qui implique qu'il encode le modèle du monde le plus large ; (2) il a un plus haut niveau d'activité métabolique que n'importe quel réseau neuronal, ce qui implique une fonctionnalité élevée¹¹ ; (3) il a un niveau soutenu d'activité, ce qui veut dire qu'il démontre à la fois une activité stable (*steady state*) et des dynamiques transitives (*transient*)¹² ; (4) il s'active durant des activités cognitives de haut niveau qui sont le propre de l'être

¹¹ En effet, une grande partie du budget énergétique du cerveau est réservée pour l'activité neuronale spontanée, une activité qui a été systématiquement associée au réseau du mode par défaut (Fox et Raichle, 2007 ; Raichle et Snyder, 2007). Carhart-Harris et Friston (2010) spéculent que l'activité spontanée dans le DMN reflète la contenance constante de l'activité endogène spontanée du cerveau, une activité qui serait principalement issue des variations homéostatiques encodées par l'intéroception (et ressenties à travers l'affect), ainsi que des motivations instinctuelles qui cherchent à les résoudre (les pulsions). Ils considèrent ce processus comme étant équivalent à celui que Freud capture à travers le concept de refoulement.

¹² La recherche a montré que le DMN n'est pas seulement impliqué dans l'autorégulation émotionnelle et la cognition autoréférentielle ou sociale, mais qu'il est aussi activé lors d'autres formes de cognition impliquant la manipulation de concepts d'un haut niveau d'abstraction (Baetens *et al.*, 2014 ; Vatansever *et al.*, 2018). De plus, des études récentes ont montré que les nœuds du DMN sont aussi impliqués dans des tâches normalement associées aux autres grands réseaux, comme les processus de contrôle exécutif impliqués dans la mémoire de travail, (Bluhm *et al.*, 2011), le rappel relié à des tâches spécifiques (Simons *et al.*, 2007), le traitement intéroceptif relié à des tâches (Gerlach *et al.*, 2011) et la cognition orientée vers un but (Spreng *et al.*, 2010). Plus précisément, ces études suggèrent que ces tâches cognitives seraient sous-tendues par l'émergence transitoire d'une corrélation positive entre l'activité de certains nœuds spécifiques du SN, du CEN et du DMN. Selon Bressler et Tognoli (2006), cette réorganisation transitoire de l'activité et de l'intégration fonctionnelle des grands réseaux en fonction du contexte pourrait fournir la flexibilité requise pour s'adapter aux demandes computationnelles changeantes du traitement cognitif. L'utilisation de la technique de la magnétoencéphalographie (qui possède une très bonne précision temporelle) suggère que cette collaboration entre des réseaux normalement anti-corrélés s'expliquerait par une alternance dynamique (*give and take*) qui s'opèrerait à l'échelle des millisecondes (Baker *et al.*, 2014 ; Cocchi *et al.*, 2013).

humain, comme la contemplation morale et la prospection (nous ajouterions la théorie de l'esprit) ; (5) les comportements liés au DMN ont une étendue temporelle très large, et portent sur des choses qui sont éloignées du moment présent ; (6) le DMN implémente une simulation du futur, ce qui constitue une entreprise pleine d'incertitude qui implique une distanciation du moment présent et un détachement relatif du flux d'entrées sensorielles qui parviennent au système nerveux ; (7) cela requiert que les dynamiques internes du système soient relativement « bruitées », pour qu'il puisse s'engager dans un mode exploratoire qui lui permet de se déplacer librement d'une idée à l'autre (Tsuda, 2009) ; réduire ce bruit (entropie) à un niveau optimal qui n'empêche pas l'exploration (l'ouverture d'esprit et la pensée créative) tout en ne permettant pas la pensée fallacieuse ou magique est selon Carhart-Harris et Friston (dans Fotopoulou *et al.*, 2012) ce qui caractériserait les processus secondaires de l'Ego.

Nous terminons notre exposé de la proposition de Carhart-Harris et Friston avec la citation suivante, dans laquelle ils résument parfaitement, leur position :

In brief, we propose that the DMN is the core of the ego. Outside of psychoanalytic parlance, it can be called 'the self.' Expressed crudely, but for the sake of explanation, it is *the thing that is always there*; the highest-level emergent property of self-organized criticality in the brain. It contains representations of our pasts, and it is the constraining edge of our desires. Its explorative cognitive style allows us to mentally simulate scenarios that are far removed from the immediate moment—but which may be essential for social cohesion and prosperity. Doing away with a dualism that protests that the ego/self cannot be reduced to the brain is a natural progression for psychoanalysis—and an agenda of the neuropsychology movement. It is the very stuff of secondary process thinking and the antithesis of a wishful primary process. (Carhart-Harris et Friston dans Fotopoulou *et al.*, 2012).

Tel que nous pouvons le voir dans la citation ci-dessus, Carhart-Harris et Friston affirment que l'Ego peut aussi être appelé le Soi (*Self*). Cependant, ces deux termes ne sont pas exactement des synonymes. Employé presque exclusivement dans le cadre de

la théorie psychanalytique et de la psychologie psychodynamique, le terme d'Ego est utilisé pour référer à une structure psychologique responsable de l'accomplissement d'un ensemble de fonctions psychologiques essentielles, et c'est à cette structure fonctionnelle que Carhart-Harris et Friston identifient l'activité du réseau du mode par défaut. Le terme de Soi a un usage beaucoup plus répandu, ce qui fait que sa signification varie selon le contexte. Dans le cadre de la philosophie de l'esprit, le Soi peut-être défini comme la conscience réflexive qu'un sujet humain a de *lui-même* en tant qu'individu (Zahavi, 2008). Autrement dit, il s'agit de ce qui est perçu ou visé lorsqu'un individu se prend lui-même pour objet de son expérience subjective. En psychologie, le Soi est souvent identifié au concept de soi (self-concept), qui correspond à une synthèse de l'ensemble des représentations descriptives et évaluatives qu'un individu entretient à propos de lui-même, et qui forment la base de son identité personnelle (Bracken, 1996 ; Hattie, 2014 ; McAdams, 2013). Bien sûr les significations philosophique et psychologique du Soi se recouvrent, étant donné que l'entretien d'un concept de soi implique forcément des activités cognitives dans lesquelles l'individu se prend lui-même comme l'objet de son expérience subjective.

Il est intéressant de noter que l'activité du DMN a été associée de manière consistante à la cognition autoréférentielle (Gusnard *et al.*, 2001 ; Molnar-Szakacs et Uddin, 2013 ; Moran *et al.*, 2013 ; Sheline *et al.*, 2009) à travers laquelle s'élabore le concept de soi. C'est ce qui a mené certains chercheurs à proposer que le DMN pourrait constituer le siège du Soi psychologique (Davey *et al.*, 2016 ; Qin et Northoff, 2011 ; Uddin *et al.*, 2007). Le fait que les fonctions cognitives associées à l'Ego psychanalytique et au Soi psychologique puissent être sous-tendues par les mêmes substrats neuronaux suggère un recouvrement des phénomènes capturés par ces deux concepts. Dans le même ordre d'idée, la grande importance accordée par Carhart-Harris et Friston (2010) aux habiletés d'autorégulation émotionnelles dans la description des fonctions du DMN est en directe continuité avec le rôle central attribué à ces dernières dans la théorie de la mentalisation, qui y sont considérées comme la fonction principale de l'Ego ou *du Soi*.

Dans la théorie de la mentalisation, le terme de Soi est privilégié à celui d'Ego, mais la signification qui lui est accordée correspond à une forme hybride de ces deux concepts qui englobe l'ensemble des phénomènes qu'ils capturent (Fonagy *et al.*, 2018). En effet, selon la théorie de la mentalisation, l'émergence du concept du Soi psychologique est une conséquence du développement et de l'exercice de l'habileté cognitive qu'elle considère comme étant le cœur des fonctions de l'Ego : *la théorie de l'esprit*.

Aussi appelée capacité de mentalisation, cette habileté qui consiste à comprendre ses états affectifs et ses comportements (ainsi que ceux des autres) en termes d'états mentaux intentionnels (ex. émotions), constituerait la fonction cognitive centrale qui sous-tendrait l'autorégulation émotionnelle et la cognition sociale. Le développement et l'utilisation du concept de Soi psychologique afin de réguler ses émotions et ses comportements résulteraient de l'exercice de cette capacité de mentalisation, qui mènerait à l'émergence d'une forme d'agentivité dans laquelle l'individu se conçoit et se comporte comme un agent cognitif sujet à des états mentaux intentionnels : autrement dit, comme un *esprit*, un *agent mental*, ou un *être psychologique*. Rappelons ici que la capacité de mentalisation a elle aussi été associée de manière systématique à l'activité du DMN (Denny *et al.*, 2012 ; Frith *et al.*, 2003 ; Hyatt *et al.*, 2015 ; Lombardo *et al.*, 2009 ; Spreng *et al.*, 2008 ; Spunt *et al.*, 2010, 2015), de telle manière qu'il n'est pas rare que ce dernier soit appelé « mentalizing network » (Barrett et Satpute, 2013 ; Hoffmann, 2020 ; Hoskinson *et al.*, 2019 ; Muscatell *et al.*, 2012).

Les remarques précédentes nous poussent à formuler l'idée suivante : bien qu'ayant été développée dans but de défendre la validité théorique de la psychanalyse freudienne, nous pensons que la proposition de Carhart-Harris et Friston d'identifier le réseau du mode par défaut à l'Ego et aux processus secondaires peut être étendue afin de tenter définir de l'espace d'échelle des phénomènes qui sous-tendent la pratique de la psychothérapie psychodynamique dans son ensemble. En effet, nous pensons que l'échelle d'organisation neurocognitive incarnée par le réseau du mode par défaut

constitue la meilleure candidate pour l'explication de la psychothérapie psychodynamique. Depuis sa découverte, plus de 3000 articles scientifiques ont été écrits afin de découvrir la nature et la fonction de ce réseau, et il a été associé de manière consistante à un grand nombre des activités cognitives qui constituent le cœur de la pratique de la psychothérapie psychodynamique telle que nous l'avons définie dans notre *explanandum* : la cognition autoréférentielle et introspective (Gusnard *et al.*, 2001 ; Molnar-Szakacs et Uddin, 2013 ; Moran *et al.*, 2013 ; Sheline *et al.*, 2009) ; la cognition sociale (Amft *et al.*, 2015 ; Li *et al.*, 2014 ; Mars *et al.*, 2012 ; Spreng et Andrews-Hanna, 2015) ; la régulation et la catégorisation des émotions (Hervé *et al.*, 2013 ; Immordino-Yang, 2016 ; Martins et Mather, 2016 ; Satpute et Lindquist, 2019 ; Wiebking *et al.*, 2011) ; la mémoire autobiographique (Hoffmann, 2020 ; Spreng *et al.*, 2008) ; la métacognition (Berkovich-Ohana *et al.*, 2012 ; Simon et Engström, 2015 ; Smigielski *et al.*, 2019 ; Valk *et al.*, 2016), etc.

Afin de tenter de défendre l'hypothèse formulée ci-dessus, il nous faut la préciser en tentant de définir comment le réseau du mode par défaut peut être considéré comme un espace d'échelle à part entière de l'organisation des systèmes neurocognitifs humains, et ce afin de répondre aux exigences du modèle ontologique de la neuroéthologie variationnelle que nous avons présenté dans notre *explanans*. Pour ce faire, nous allons nous baser sur le travail réalisé par Connolly et van Deventer (2017), qui affirment que c'est le cadre conceptuel de la théorie générale des systèmes qui fournit le contexte critique permettant la meilleure compréhension de la manière dont le FEP peut-être utilisé pour expliquer les phénomènes décrits par la théorie psychanalytique¹³. Leur position est très bien résumée par la citation suivante :

¹³ C'est à partir de la position de Connolly et van Deventer (2017) que nous avons choisi d'adopter la théorie générale des systèmes comme cadre conceptuel principal de notre mémoire.

... the ego must be understood as self-organizing (Grobbelaar, 1989), and the specific nature of that self-organizing process is itself emergent from the FEP (Connolly, 2016). The ego is understood here as an associative structure occupying the higher levels of organization of the generative model, that comes to influence lower levels of the hierarchy. As such, it develops Markov characteristics that mean it (the ego) must be viewed as effectively self-organizing—and potentially self-evidencing as described by Hohwy (2016). Psychoanalysis is proposed here as being essentially that science of the self-organization of the ego that describes its relative inertia and resistance to change, while also describing the unique principles of organization that operate at this level.

La proposition explicitée dans cette citation est relativement simple, et consiste à considérer l'Ego comme un niveau d'auto-organisation spécifique des systèmes neurocognitifs humains, qui émerge de l'activité des échelles supérieures de la hiérarchie prédictive incarnée par le cerveau. Tel que prescrit par la théorie générale des systèmes, ce niveau organisation est défini par un ensemble de processus interreliés, qui d'après la théorie autopoïétique du vivant, travaillent à l'autoproduction et à l'autodistinction de l'échelle systémique qu'il incarne. Selon le FEP, ces processus travaillent plus spécifiquement à la minimisation de l'énergie libre des états internes de cette échelle systémique particulière, qui doivent être délimités des états externes à travers l'émergence d'une couverture de Markov qui lui est propre (ou plusieurs). Cependant, d'après la neuroéthologie variationnelle, l'impératif de minimisation de l'énergie libre est implémenté à travers des mécanismes hétérogènes à travers les différentes échelles d'organisation des systèmes vivants. Selon Connolly et van Deventer (2017), dans le cas de l'échelle d'organisation incarnée par l'Ego, ces mécanismes correspondent aux processus mentaux qui sont étudiés par la psychanalyse (et qui sont selon nous impliqués dans la psychothérapie psychodynamique). C'est la raison pour laquelle ils affirment que la théorie psychanalytique correspondrait à la discipline scientifique dédiée à l'étude des principes d'organisation de l'Ego. Cette affirmation est selon nous précipitée, car comme nous l'avons vu précédemment, la

théorie psychanalytique n'a pas été développée à l'aide de l'application de la méthode scientifique.

En effet, et tel que nous l'avons vu dans notre introduction, l'activité d'un niveau d'auto-organisation ne peut pas excéder les limites d'action potentielles des échelles d'organisation inférieures dont il émerge. Si nous suivons la directive de régulation épistémique pour laquelle le FEP a été développé, alors l'étude des principes d'organisation de l'Ego devrait être régulée par les connaissances à propos des échelles d'organisation proximales dont il émerge, et qui correspondent au domaine d'étude des neurosciences cognitives (et des autres sciences de l'esprit contemporaines comme l'anthropologie cognitive, la philosophie de l'esprit, etc.). C'est la raison pour laquelle nous pensons, et cela est en directe continuité avec le projet de la Neuropsychanalyse, que la théorie psychanalytique devrait être remplacée par un modèle théorique basé l'intégration de la psychologie psychodynamique au sein des neurosciences cognitives contemporaines. C'est précisément avec cet objectif en tête que dans un ouvrage collectif de grande ampleur, Fotopoulou *et al.* (2012) ont proposé et entamé le développement d'un nouveau sous-champ disciplinaire des neurosciences cognitives qu'ils ont choisi de nommer *les neurosciences psychodynamiques*. C'est dans le projet incarné par ce nouveau champ d'études que s'inscrivent les propositions que nous allons formuler dans cette dernière partie de notre mémoire.

L'hypothèse que nous souhaitons retenir du travail réalisé dans cette section est la suivante : il est possible de comprendre les activités cognitives impliquées dans la pratique de la psychothérapie comme les processus qui définissent un niveau d'auto-organisation particulier des systèmes neurocognitifs humains, et qui correspond à un sous-ensemble de l'activité des échelles supérieures de la hiérarchie prédictive incarnée par le cerveau. Si nous avons choisi de nous concentrer sur le réseau du mode par défaut, c'est parce qu'il correspond au substrat neuronal qui présente pour l'instant le plus de

potentiel pour l'explication de la psychothérapie psychodynamique¹⁴. Dans ce qui suit, nous allons tenter de décrire plus précisément quels pourraient être les processus qui définissent ce niveau d'organisation, de spécifier le processus développemental à travers lequel il en vient à émerger, et comment la pratique de la psychothérapie psychodynamique pourrait venir le modifier. Pour ce faire, nous allons prendre le parti de la théorie de la mentalisation, et considérer les habiletés d'autorégulation émotionnelle et la théorie de l'esprit comme les fonctions principales qui sont implémentées par les processus qui définissent ce niveau d'organisation. Pour référer à ce niveau d'organisation, nous avons choisi de privilégier le concept de Soi dans le sens élargi qui lui est attribué par la théorie de la mentalisation, car il couvre selon nous de manière plus exhaustive les phénomènes impliqués dans la psychothérapie psychodynamique, en plus de jouir d'un usage plus répandu.

3.1.2 L'émergence du Soi à travers le couplage structurel d'un individu avec l'écologie socioculturelle de l'espèce humaine

Pour commencer, nous devons nous former une compréhension minimale du processus à travers lequel les systèmes vivants en viennent à développer de nouveaux niveaux d'organisation. Pour ce faire, nous allons nous baser sur la réflexion amenée par Ramstead et al. (2017), qui partent du FEP pour avancer la thèse suivante : un système vivant ne fait pas que performer une modélisation de son environnement, il est un

¹⁴ Cependant, nous ne sommes pas fermés à la possibilité que les activités cognitives sous-tendues par la psychothérapie psychodynamique puissent impliquer des substrats neuronaux plus (ou moins) étendus que le réseau du mode par défaut.

modèle de son monde. Cette thèse dérive directement de la dépendance d'un organisme à sa niche écologique pour lui fournir les ressources nécessaires au processus autopoïétique qui le définit. L'environnement d'un organisme forme lui-même un système écologique complexe qui possède ses propres dynamiques d'organisation, définies par les interactions entre les composantes biotiques et abiotiques qui le composent. La théorie générale des systèmes et la neuroéthologie variationnelle insistent sur le fait que les processus qui définissent l'organisation des systèmes physiques et vivants s'étalent sur différentes échelles spatiotemporelles récursivement imbriquées les unes dans les autres. Cette perspective nous permet de réaliser le fait que le couplage structurel qui unit un système vivant et le système écologique dans lequel il s'inscrit s'opère lui aussi à différentes échelles spatiotemporelles.

En effet, nous pouvons observer un parallélisme entre les échelles d'organisation des systèmes vivants et celles de leur niche écologique. Le processus d'adaptation d'un système vivant à chacune des dimensions de sa niche écologique mène nécessairement à l'émergence d'une échelle spécifique de son organisation qui résulte du couplage structurel avec une échelle environnementale particulière. D'après le *Skillful Intentionality Framework* (SIF; (Bruineberg et Rietveld, 2014), ce couplage structurel passe par le développement d'habiletés qui permettent à l'organisme de percevoir et de s'engager dans un paysage d'affordances élargi, des opportunités d'interactions adaptatives avec cette échelle particulière de sa niche écologique. L'exercice de ces habiletés va maintenant former partie intégrante des processus qui définissent l'organisation de ce système vivant, et les dynamiques d'organisation qu'elles incarnent vont exercer une influence déterminante sur l'ensemble du système en imposant de nouvelles contraintes sur l'activité des échelles d'organisation inférieures dont elles émergent. Ce processus de complexification des dynamiques d'organisation des systèmes vivants peut être conceptualisé comme l'émergence de nouvelles couvertures de Markov qui vont venir installer la clôture opérationnelle nécessaire au maintien de l'indépendance conditionnelle et de l'autonomie du système

à de nouvelles échelles spatiotemporelles (et informationnelles) de son couplage structurel avec sa niche écologique (Kirchhoff *et al.*, 2018 ; Palacios *et al.*, 2020).

La réflexion amenée ici nous permet de dresser le constat suivant quant à l'hypothèse formulée précédemment : si l'Ego ou le Soi correspondent bel et bien à un niveau à part entière de l'organisation des systèmes neurocognitifs humains, alors ce dernier doit émerger du couplage structurel avec un espace d'échelle particulier de la niche écologique de l'espèce humaine. Autrement dit, il nous faut tenter de répondre à la question suivante : de quelle échelle de l'interaction entre un organisme humain et sa niche écologique (du métasystème cerveau-corps-environnement) le Soi est-il le modèle ? L'une des thèses fondamentales de la psychologie psychodynamique nous offre une piste de réponse à cette question. Cette dernière consiste à affirmer que le fonctionnement cognitif et émotionnel typiquement humain serait le résultat d'un conditionnement social déterminant qui s'opère durant l'ontogénie à travers les relations interpersonnelles entre un enfant et ses donneurs de soins primaires. Cette thèse s'étend dans la conception biosociale du développement cognitif et émotionnel (et de l'émergence du soi psychologique) défendue par la théorie de la mentalisation, et de nombreuses variantes de cette thèse ont été développées au sein des neurosciences cognitives et affectives dans les dernières décennies (pour une revue extensive, voir Schore, 2016). Ces théories partent toutes du constat de l'importance critique de l'organisation sociale de l'espèce humaine dans le développement des capacités cognitives qui la caractérisent.

En effet, l'espèce humaine est une espèce sociale (Hinde, 1974). De plus, et tel que mentionné dans notre introduction à la théorie de l'attachement, l'espèce humaine est une espèce *altricielle* (Dunsworth *et al.*, 2012 ; Roberts et Thorpe, 2014 ; Zeveloff et Boyce, 1982), ce qui veut dire que le nouveau-né humain est typiquement prématuré, incompetent, et vulnérable, et qu'il est dépendant de l'aide de son environnement social pour assouvir ses besoins biologiques fondamentaux, et ce pendant une période

développementale particulièrement longue (Sleigh, 2011). C'est autour de ce constat que sont organisés beaucoup des développements récents en anthropologie cognitive, qui affirment que les spécificités du fonctionnement cognitif humain émergent d'un processus ontogénétique complexe au sein duquel les interactions entre l'individu et son environnement social, et en particulier les relations avec ses donneurs de soins primaires, jouent le rôle principal (Hrdy, 2009, 2017 ; Tomasello, 1999, 2014, 2019 ; Tomasello et Carpenter, 2007 ; Tomasello et Gonzalez-Cabrera, 2017). Il a été proposé que la période ontogénétique étendue de l'espèce humaine serait accompagnée d'un maintien prolongé de la plasticité neuronale qui est associée à la période initiale du développement du système nerveux central des mammifères, ce qui pourrait expliquer en partie la flexibilité et la créativité qui caractérisent la cognition humaine (Boesch, 2020 ; Bufill *et al.*, 2011 ; Ebbesson, 1980 ; Johnson, 1999 ; McKinnon et Silverman, 2005 ; Plebe et Mazzone, 2016 ; Somel *et al.*, 2009 ; Vartanian *et al.*, 2013). Pris ensemble, ces travaux étayent l'hypothèse selon laquelle l'environnement social pourrait jouer un rôle dominant dans le développement du cerveau et du fonctionnement cognitif des être humains.

Dans un article récent intitulé *Growing a Social Brain*, Atzil *et al.* (2018) formulent une variante de cette hypothèse dans le langage du FEP et des approches prédictives du cerveau humain. Tout d'abord, ils proposent que les animaux sociaux puissent être définis comme des animaux qui ne peuvent pas survivre seuls et qui dépendent des membres de leur groupe pour réaliser leur propre allostasie (c.-à-d.. se maintenir dans l'ensemble d'états délimités par leurs phénotypes afin de minimiser leur propre énergie libre). Ils affirment ensuite que cette simple contrainte évolutionnaire suffise pour faire de l'environnement social la force causale la plus déterminante pour le développement du cerveau et de la cognition humaine. Plus précisément, ils affirment qu'en plus de simplement permettre la réalisation de l'allostasie de l'enfant en développement, les interactions de modulation homéostatique avec ses donneurs de soins primaires profitent de la motivation intrinsèque associée à ce processus pour inciter

l'apprentissage d'habiletés cognitives essentielles au fonctionnement psychologique humain. Parmi ces dernières, celles qui nous intéressent le plus sont les habiletés d'autorégulation émotionnelles, et plus particulièrement la fonction cognitive qui les sous-tendrait : la théorie de l'esprit, ou capacité de mentalisation. Tel que nous l'avons présenté dans notre *explanandum*, la théorie de la mentalisation possède déjà un modèle du processus par le biais duquel la capacité mentalisation est acquise à travers les relations interpersonnelles durant le développement. Afin de montrer comment ce processus ontogénétique mène à l'émergence du Soi psychologique à travers le développement d'un nouveau niveau d'organisation des systèmes neurocognitifs humains, il nous reformuler ce modèle dans les termes du cadre conceptuel avancé dans notre *explanans*.

Pour ce faire, nous prendrons comme point de départ la position développée par Fotopoulou et Tsakiris (2017), qui proposent une définition de l'échelle la plus fondamentale du Soi, le soi minimal corporel et affectif, dans le cadre du FEP et des approches du traitement prédictif. Le modèle qu'ils développent repose sur une extension radicale du concept de mentalisation, la « mentalisation incarnée », qu'ils définissent comme un processus d'inférence par lequel les signaux intéroceptifs, sensorimoteurs et multi sensoriels primaires sont progressivement intégrés et schématisés pour former des modèles prédictifs de nos *états corporels* (par opposition aux états mentaux) dans des environnements donnés. Selon eux, ce processus mène à l'inférence du *Soi*, une hypothèse qui est théorisée comme l'explication la plus parcimonieuse et adéquate des entrées sensorielles. Ainsi, le *Soi* correspondrait à un modèle prédictif qui explique les régularités des domaines intéroceptifs et proprioceptifs, ainsi que les contingences sensorimotrices qui les lient au domaine de l'extéroception, comme résultant d'une seule et même entité causale cohérente : l'organisme lui-même. L'émergence de ce modèle prédictif se manifesterait sur le plan phénoménologique par la constitution du sujet de l'expérience perceptuelle, de la conscience affective, et des états subjectifs fondamentaux comme l'expérience

cohérente et continue de posséder un corps qui est le nôtre, et que nous pouvons contrôler de manière volontaire.

De plus, Fotopoulou et Tsakiris (2017), avancent une perspective réductionniste et mécaniste de la manière dont les interactions sociales, dans l'ontogénie et au-delà, mènent à la constitution du soi minimal corporel et affectif. Puisque les enfants humains naissent avec un système moteur immature, ce sont les donneurs de soins primaires (les parents en général) qui, par leurs propres inférences actives, assurent la réalisation de l'allostasie de leurs enfants. Les réflexes autonomes (ex. pleurer et donner des coups de pied) de l'enfant sont conceptualisés par la théorie de la mentalisation incarnée comme des moyens de communiquer *l'inférence perceptuelle* (intéroceptive) au donneur de soin, qui doit l'interpréter et répondre avec *l'inférence active* correspondant au besoin perçu chez l'enfant (ex. le nourrir ou le prendre dans ses bras). Ce faisant, les actes de soins fournis par le donneur de soins démontrent à l'enfant les inférences actives qui peuvent être mises en place afin de réaliser son allostasie. Ainsi, il est possible de définir le soi minimal corporel et affectif comme un niveau d'auto-organisation neurocognitif qui incarne un modèle de l'état homéostatique du corps de l'organisme et de sa situation dans son environnement physique. Dans le cas d'un organisme humain, cette échelle de son environnement est composée en grande partie d'autres organismes humains qui ont un pouvoir causal déterminant sur la réalisation de son allostasie. Cela explique pourquoi le modèle qu'il incarne est en grande partie généré à partir des informations provenant des interactions avec ces autres organismes.

Remarquons ici que le développement d'un soi minimal corporel et affectif est probablement quelque chose qu'une grande majorité des animaux vertébrés partagent. Étant donné la dépendance extrême et prolongée de sa progéniture, la particularité de l'espèce humaine réside probablement dans l'influence accrue que les interactions sociales jouent dans son développement, ce qui pourrait avoir des conséquences

durables sur le fonctionnement cognitif d'un individu adulte. Par opposition au soi minimal corporel et affectif, le développement d'un soi psychologique, et des habiletés cognitives qui le sous-tendent, semblent être le propre de l'être humain. Dans la section suivante, nous allons prendre le modèle de la mentalisation incarnée comme point de départ afin de tenter de monter dans les échelles et de progresser du soi corporel et affectif vers le soi émotionnel et psychologique décrit originellement par la théorie de la mentalisation, et adressé par la psychothérapie psychodynamique.

Afin d'élaborer une explication de la manière dont la capacité de mentalisation et le fonctionnement émotionnel d'un individu sont acquis à travers les relations interpersonnelles durant le développement — et peuvent être modifiés à travers la psychothérapie — il nous faut procéder en deux étapes : (1) nous devons reformuler la capacité de mentalisation dans les termes du FEP et des approches prédictives du cerveau humain ; et (2) il nous faut proposer une explication des mécanismes par le biais desquels cette capacité est transmise à travers les interactions sociales.

3.1.3 La mentalisation et le Soi psychologique selon les approches prédictives du cerveau humain

Comme nous l'avons vu précédemment, l'échelle d'organisation neurocognitive incarnée par le réseau du mode par défaut a été associée à la fois au Soi psychologique, aux capacités d'autorégulation émotionnelles, et à la capacité de mentalisation, ce qui est en soi un appui substantiel pour l'hypothèse (de la théorie de la mentalisation, voir Fonagy *et al.*, 2018) selon laquelle ces phénomènes psychologiques seraient fondamentalement interdépendants. Barrett et Satpute (2013) proposent que le réseau du mode par défaut, qu'ils appellent « mentalizing network », puisse performer une

seule et même fonction à travers les domaines sociaux, affectifs, et cognitifs du traitement de l'information implémenté par le cerveau. Ils font l'hypothèse que les nœuds du réseau de la mentalisation interagissent pour créer un système d'inférence qui génère des *conceptualisations situées* : des simulations multimodales qui sont ancrées dans un contexte particulier, afin de donner un sens aux entrées sensorielles, et de guider des plans d'actions spécifiques. Leur hypothèse est que le réseau de la mentalisation dirige la recombinaison associative d'informations stockées dans la mémoire (déclarative et procédurale) pour créer non seulement des instances de cognition, mais aussi des instances d'émotions et de perceptions, incluant notamment la perception de *personnes* et d'événements dans le monde. Autrement dit, l'activité du réseau du mode par défaut serait responsable de la capacité qu'à un individu de se percevoir lui-même et les autres organismes humains qui l'entourent comme des « personnes », et de leur attribuer des états particuliers comme des émotions, des intentions, ou des traits de personnalité afin d'expliquer et de prédire leurs comportements.

Selon nous, il est évident que l'ensemble de ces concepts appartiennent à la catégorie générale de ce que l'on appelle des états mentaux, et que l'habileté cognitive qui permet de les implémenter correspond exactement à ce que Fonagy *et al.* (2018) appellent la capacité de mentalisation. En d'autres termes, nous pensons qu'il est possible de concevoir la mentalisation comme une activité de conceptualisation qui porte spécifiquement sur la modélisation des dynamiques d'organisation cognitive des êtres humains. Cette hypothèse est très similaire à celle avancée par Friston et Frith (2015), qui conçoivent la théorie de l'esprit comme un modèle prédictif qui dérive directement du processus d'inférence à travers lequel le cerveau tente de minimiser l'énergie libre de l'organisme auquel il appartient. Si l'on récapitule le modèle du cerveau humain proposé par les approches prédictives, nous avons une machine à inférence qui doit modéliser les états externes du monde qui causent les entrées sensorielles qu'elle reçoit. Or, dans le cas d'un organisme humain, une grande partie des forces causales qui

expliquent ces sensations correspondent aux actions des autres organismes humains. Puisque ces actions sont principalement causées par l'activité du système nerveux central de ces organismes, alors cela implique que pour les prédire efficacement, le cerveau d'un individu doit modéliser l'activité du cerveau des autres individus, et cela de manière réciproque. Friston et Frith (2015) affirment que la notion même de théorie de l'esprit implique directement un processus d'inférence, dans le sens où les théories font des prédictions qui doivent être testées par rapport à des données (sensorielles ici). Ils poursuivent en proposant que, dans le cadre des approches du traitement prédictif, la théorie de l'esprit puisse être comprise comme l'inférence de *l'état d'esprit* dans lequel une personne devrait se trouver pour produire les mêmes conséquences sensorielles.

Continuant sur cette lancée, Moutoussis *et al.* (2014) ont proposé de concevoir la mentalisation comme une forme *d'inférence active interpersonnelle* dont le but est de prédire les conséquences des échanges interpersonnels, et ce afin de favoriser l'adaptation de l'organisme à son environnement social. En appui à cela, la proposition de Friston et Frith (2015) consiste à affirmer que la théorie de l'esprit est le modèle le plus efficace pour prédire les dynamiques comportementales de l'individu et des autres agents qui composent sa niche écologique, et que ceux-ci constituent par conséquent l'explication la plus probable des causes des signaux sensoriels qui parviennent au système nerveux. En directe continuité avec la théorie de la mentalisation, Friston et Frith (2015) affirment que les modèles internes générés par un individu pour inférer son propre état (inférence perceptuelle autoréférentielle) et son propre comportement (inférences actives) sont les mêmes qui vont être utilisés pour inférer les croyances (inférences perceptuelles) et les intentions (plans d'inférences actives) d'autrui, du moins dans le cas où les individus possèdent un modèle génératif suffisamment similaire. Autrement dit, ils suggèrent que les croyances préalables (*prior beliefs*) à propos des causes du comportement d'un être humain (et de ses conséquences) ne sont pas nécessairement liées à un agent particulier, mais qu'elles sont utilisées pour

reconnaitre des stéréotypes comportementaux (et cognitifs) qui sont parfois générés par nous-mêmes, et parfois par autrui. D'après Fonagy et Luyten (2012), cette hypothèse est appuyée par la recherche en neuroimagerie (Dimaggio *et al.*, 2008 ; Lieberman, 2007 ; Lombardo *et al.*, 2009 ; Uddin *et al.*, 2007), qui suggère que la capacité de se représenter les états mentaux d'autrui est sous-tendue par les mêmes systèmes cérébraux que la capacité d'identifier ses propres pensées et sentiments.

En synthétisant l'ensemble des propositions présentées ci-dessus au modèle des émotions proposé par Barrett (2016), nous pensons pouvoir formuler l'hypothèse selon laquelle états mentaux peuvent être considérés comme une catégorie particulière de concepts neurocognitifs d'un niveau d'abstraction élevé, qui expliquent l'état et le comportement des êtres humains en faisant l'hypothèse qu'ils sont déterminés par un type particulier de causes distales : les processus qui définissent l'échelle cognitive (ou mentale) de leur organisation. En d'autres termes, nous pensons que dans le cadre des approches prédictives de la cognition humaine, la capacité de mentalisation peut être définie comme une habileté de *conceptualisation métacognitive* : une activité de modélisation prédictive des processus cognitifs qui gouvernent le comportement des être humains. Ici, nous utilisons le concept de métacognition en son sens le plus basique, c'est-à-dire pour référer à toute forme d'activité cognitive qui porte sur les processus cognitifs en eux-mêmes et sur leurs produits ou leurs manifestations (Brown, 1987 ; Flavell, 1976 ; Metcalfe *et al.*, 1994), et qui peut s'accompagner d'une activité de régulation de ces processus cognitifs (Schraw, 1998). En ce sens, la mentalisation est une forme de métacognition en ce qu'elle consiste en une activité cognitive qui porte elle-même sur la modélisation de l'activité cognitive d'un être humain. Il en est de même pour l'activité de conceptualisation émotionnelle décrite par Barrett (2016), qui incarne une forme de modélisation prédictive visant à expliquer et à réguler les processus neurocognitifs qui sont responsables de l'expérience affective et des comportements de l'individu. Comme nous le verrons par la suite, le type de conceptualisation métacognitive que nous identifions à la capacité de mentalisation

comprend selon nous à la fois des formes de cognition explicites et réflexives, mais aussi et majoritairement des formes de cognition implicites et automatiques qui appartiennent au domaine de l'inconscient cognitif. Pour le moment, essayons d'appliquer notre nouvelle compréhension de la capacité de mentalisation afin de voir comment elle peut être responsable de l'émergence du soi psychologique.

Pour ce faire nous allons suivre Moutoussis *et al.* (2014), qui proposent que les représentations du Soi (psychologique) correspondent aux modèles génératifs issus du processus d'inférence active (mentalisation) à travers lequel un individu infère ses propres comportements et leurs conséquences dans le contexte des échanges interpersonnels, qui constituent l'une des dimensions les plus importantes de son interaction avec sa niche écologique. Selon eux, ce qu'ils appellent le « soi interpersonnel », est le résultat d'un processus d'inférence probabiliste qui vise à minimiser l'incertitude reliée au échanges sociaux. Autrement dit, le soi est conçu ici comme un modèle prédictif du comportement de l'individu dans l'exercice de son agentivité sociale, un modèle basé sur l'inférence que les causes distales qui sont à l'origine de ce comportement correspondent aux états qui définissent l'échelle cognitive de son organisation. Plus précisément, et en concordance avec la notion de concept de soi développée en psychologie, Moutoussis *et al.* (2014) suggèrent que le processus d'inférence qui mène aux représentations de soi joue une fonction normative : prédire et optimiser les conséquences probables des interactions sociales.

En effet, dans le cadre des approches prédictives, la distinction entre le concept psychanalytique d'Ego et le concept psychologique de Soi, correspondant respectivement à un ensemble de fonctions d'autorégulation et à un ensemble de représentations qu'un individu entretient à propos de lui-même, se retrouve dissoute. Lorsque le Soi est conçu comme un modèle prédictif de l'individu en tant qu'agent sociocognitif, alors les prédictions qu'il génère implémentent à la fois des inférences perceptuelles qui décrivent et expliquent son comportement (représentations de soi), et

des inférences actives qui régulent ce dernier afin d'implémenter un couplage structurel efficace avec son environnement social (fonctions d'autorégulation). Si l'on formule cette idée dans le langage conceptuel développé par Barrett (2016), cela revient à dire que dans le cadre des approches prédictives, le soi et le concept de soi sont une seule et même chose. Autrement dit, le concept neurocognitif de soi correspond à un modèle prédictif qui représente *et* régule les comportements et les états affectifs de l'individu¹⁵.

Tout comme c'était le cas pour le soi minimal corporel et affectif, le postulat d'un soi, psychologique cette fois-ci, est l'explication la plus probable des informations qui parviennent à ce niveau d'organisation. Ces deux concepts sont issus d'un processus de modélisation de l'état de l'organisme lui-même à différentes échelles de son organisation : une activité qui peut être considérée selon nous comme une forme « d'auto-modélisation », et qui peut être comprise comme une extension du phénomène de « self-evidencing » proposé par Hohwy (2016). Le soi minimal corporel et affectif intègre les signaux intéroceptifs, sensorimoteurs, et multi sensoriels primaires pour former un modèle prédictif de l'état homéostatique du corps de l'organisme dans le cadre du couplage structurel avec son environnement biophysique. Le soi psychologique incarne quant à lui un modèle de l'état et de l'activité de l'organisme à une échelle plus élevée de son organisation. Il intègre les informations provenant des échelles intermédiaires de la hiérarchie cérébrale pour former un modèle prédictif des dynamiques de l'activité affective, cognitive et comportementale (c.-à-d.

¹⁵ Pour une théorie plus exhaustive et précise du Soi formulant des hypothèses similaires à celles proposées dans cette section, voir le « self-model of subjectivity » développé par Thomas Metzinger (2004)

émotionnelle¹⁶) de l'organisme qui émergent du couplage structurel avec son éconiche socioculturelle.

En intégrant la perspective formulée ci-dessus avec les propositions exposées dans la première section de ce chapitre (Carhart-Harris et Friston, 2010 ; Connolly et van Deventer, 2017), nous pensons qu'il est possible de considérer le modèle prédictif correspondant au concept de Soi psychologique comme la propriété d'une échelle d'auto-organisation spécifique du fonctionnement neurocognitif d'un organisme humain, qui émerge du couplage structurel avec son éconiche socioculturelle, et dont le rôle est de réguler (contraindre) l'activité affective, comportementale, et cognitive de l'organisme. Cette fonction serait accomplie à travers l'exercice de la capacité mentalisation, une habileté de conceptualisation métacognitive qui serait essentielle au contrôle exécutif de l'activité neurocognitive de l'organisme nécessaire pour son adaptation à un environnement socioculturel complexe. Il nous reste maintenant à voir comment cette fonction pourrait être transmise à travers les relations interpersonnelles, que ce soit durant le développement ou dans le cadre de la psychothérapie.

3.1.4 Un modèle de la communication humaine selon le principe de l'énergie libre et approches prédictives du cerveau humain

¹⁶ Dans le cadre de la théorie de Barrett (2016), les émotions correspondent à des concepts neurocognitifs qui minimisent les erreurs de prédictions reliées à l'activité affective, comportementale, et cognitive de l'organisme. Le fait que cette définition s'applique aussi au concept de soi suggère que, dans le contexte de cette théorie, ce dernier puisse être considéré comme un (méta)concept émotionnel.

Afin de fournir une explication du rôle central de l'environnement social dans l'acquisition de la capacité de mentalisation et le développement du concept neurocognitif de soi psychologique, nous allons nous baser sur un modèle de la communication humaine selon le FEP et les approches prédictives proposé par Friston et Frith (2015), et subséquemment retravaillé par Vasil *et al.* (2020). Ce modèle repose en grande partie sur le cadre conceptuel et théorique développé par l'anthropologue Michael Tomasello, que nous allons présenter brièvement ici.

À travers des décennies d'étude comparative du développement cognitif des humains et des grands singes, qui sont ses plus proches parents du point de vue phylogénétique, Tomasello et ses collègues ont développé une théorie qui affirme que l'unicité de l'espèce humaine réside dans la dimension sociale de son fonctionnement cognitif (Tomasello, 2014 ; Tomasello et Gonzalez-Cabrera, 2017 ; Tomasello, 2019). Plus précisément, Tomasello (1999) affirme que l'ensemble des habiletés uniques qui composent la forme de vie de l'espèce humaine reposent sur le développement de la culture. Dans le contexte de l'anthropologie cognitive contemporaine, la culture est définie comme toute forme d'information capable d'affecter le comportement des individus, et qu'ils acquièrent via d'autres membres de leur espèce à travers l'enseignement, l'imitation, ou d'autres formes de transmission sociale (Richerson et Boyd, 2005). La théorie dominante dans ce contexte est celle de l'évolution culturelle, qui est une théorie évolutionnaire selon laquelle une grande partie des facultés cognitives des êtres humains sont apparues, non pas par la modification de leur patrimoine génétique à travers le mécanisme biologique sélection naturelle, mais par un processus de changement social de modification de l'information culturelle à travers l'histoire des sociétés humaines (pour une formulation contemporaine de cette thèse, voir Heyes, 2018).

Tomasello et ses collaborateurs affirment que le fondement de la culture réside en un ensemble d'habiletés sociocognitives et sociomotivationnelles qu'ils regroupent sous

le concept de « l'intentionnalité partagée » (Tomasello & Carpenter, 2007). L'intentionnalité partagée, parfois appelée l'intentionnalité du « nous », réfère aux interactions collaboratives dans lesquelles les participants partagent *les mêmes états psychologiques*. À la base de cette intentionnalité partagée réside l'attention jointe, une capacité innée unique à l'espèce humaine (Moore et Dunham, 1995)¹⁷. Alors que les chimpanzés sont simplement capables de s'imaginer la perspective perceptuelle d'un autre individu, les enfants humains vont plus loin que cela. Avant même leur premier anniversaire, ils ne font pas que de suivre le regard des autres vers des cibles externes, et ne cherchent pas seulement à voir ce que les autres voient, ils essaient aussi de *partager leur attention* avec les autres (Tomasello et Carpenter, 2007). De manière importante, l'attention jointe ne correspond pas seulement à deux personnes expérimentant la même chose au même moment, mais il s'agit plutôt de deux personnes qui expérimentent la même chose au même moment et qui savent toutes les deux que c'est ce qu'ils sont en train de faire (Moore et Dunham, 1995). Il s'agit véritablement d'un partage intersubjectif, et cela est critique, car cela crée un espace partagé, un terrain d'entente psychologique qui est la base d'une grande partie des comportements humains allant des activités collaboratives avec des buts partagés jusqu'à la communication coopérative (Seemann, 2011). Cette capacité de « lire dans l'esprit des autres » (mentalisation) et de partager des expériences avec autrui (intentionnalité partagée) structurerait la vie et le monde des êtres humains dont la survie dépend grandement de l'intégration au sein d'un groupe social (Tomasello et Carpenter, 2007).

¹⁷ C'est sur la base du concept d'attention jointe que s'est développé le modèle des affordances culturelles (Ramstead *et al.*, 2016) présenté au chapitre précédent, et que nous allons utiliser dans ce qui suit pour expliquer l'acquisition développementale de la capacité de mentalisation, et sa modification par le biais de la psychothérapie.

Le modèle de la théorie de l'esprit développé par Friston et Frith (2015) s'accompagne d'un modèle du phénomène de la communication humaine selon les approches prédictives. Ce modèle suggère que le phénomène d'intentionnalité partagée résulterait d'un processus de synchronisation des dynamiques d'auto-organisation neurocognitives entre deux (ou plus) individus. Plus précisément, Friston et Frith proposent que lorsque plusieurs systèmes prédictifs qui présentent une organisation similaire (comme des cerveaux humains) sont couplés les uns aux autres, ce qui veut dire qu'ils pratiquent l'inférence active dans leur interaction afin de se prédire mutuellement, cela mène nécessairement à l'émergence d'un phénomène de synchronisation des modèles prédictifs qu'ils incarnent, et à la constitution collaborative d'un modèle génératif commun du phénotype et de la niche écologique qu'ils partagent. Dans un article publié cette année, Vasil *et al.* (2020) récupèrent et améliorent le modèle de Friston et Frith (2015) afin de proposer une explication de la communication humaine comme une forme *d'inférence active*. Leur modèle va constituer la base de notre tentative d'explication du fonctionnement de la psychothérapie psychodynamique.

Vasil *et al.* (2020) partent d'une définition de la communication coopérative comme d'une forme de communication intentionnelle utilisée par les membres d'une même espèce afin d'aligner leurs états mentaux dans le but de réaliser un objectif commun, et qui correspond souvent, du moins à l'origine, à l'assouvissement des besoins biologiques qu'ils partagent (Tomasello, 2010). À cela, ils ajoutent une précision du concept d'inférence active¹⁸ comme du processus à travers lequel chaque organisme

¹⁸ Dans les dernières années, le concept d'inférence active a subi une inflation qui l'a mené à devenir la pierre angulaire des modèles de la cognition développés dans le cadre du FEP (Brown *et al.*, 2013 ; Constant *et al.*, 2019 ; Friston *et al.*, 2017 ; Kirchoff *et al.*, 2018 ; Pezzulo *et al.*, 2015, 2018 ; Seth et Friston, 2016). Cela est du, au moins en partie, à l'intégration de l'emphase pragmatiste sur l'action adaptative prônée par les approches énaactives de la cognition (Ramstead *et al.*, 2020).

optimise son modèle interne du monde à travers des *cycles d'action-perception* qui minimisent l'énergie libre de ses états internes. De plus, l'approche développée par Vasil et al. repose sur l'hypothèse de l'existence de *priors*¹⁹ *adaptatifs*, définis comme un ensemble de croyances²⁰ héréditaires partagées par les membres d'une même espèce, et qui ont été sculptés par la sélection naturelle et/ou l'évolution culturelle afin de guider les cycles d'action-perception vers des conséquences adaptatives et non surprenantes (Badcock *et al.*, 2019b, 2019a). De tels *priors* dépendent d'un héritage génétique, épigénétique, et/ou culturel, et incorporent souvent des *priors empiriques* acquis à travers l'expérience pour permettre l'adaptation à l'environnement local (Badcock *et al.*, 2019b). Selon Vasil *et al.* (2020), les *priors adaptatifs* contraignent l'ensemble des modèles prédictifs (c.-à-d. des concepts neurocognitifs) qui peuvent être appris durant l'ontogénie afin de favoriser l'action adaptative des organismes individuels au sein de leurs niches socioculturelles locales.

Vasil *et al.* (2020) partent du constat du caractère universel de l'attention jointe (Moore et Dunham, 1995 ; Tomasello, 2010 ; Tomasello et Carpenter, 2007) pour proposer que

¹⁹ Ici, nous avons choisi d'utiliser le terme anglais de « prior », faute de l'existence d'une traduction française satisfaisante. Dans le cadre des approches prédictives du cerveau humain, un prior correspond à un modèle prédictif généré à partir d'informations collectées préalablement, et qui encode la probabilité (ou l'incertitude) qu'un (ou plusieurs) événement(s) sensoriel(s) donné(s) se produisent avant que de nouvelles données sensorielles ne soient collectées. Comme nous l'avons vu dans la section 2.2.2, ces *priors* ne sont pas seulement descriptifs, mais incluent un ensemble de valeurs existentielles qui encodent les normes phénotypiques au sein desquelles un organisme doit se maintenir s'il veut rester en vie. Le concept de prior adaptatif vise à capturer l'ensemble de ces *priors*, y compris ceux qui ne sont pas nécessairement innés.

²⁰ Le terme de croyance est ici employé dans le sens technique qui lui est octroyé dans le cadre des probabilités bayésiennes. Dans ce contexte, il n'est pas à comprendre dans le sens d'une croyance propositionnelle, mais comme un modèle prédictif qui encode la probabilité qu'un événement se produise ou non. Dans ce contexte, les croyances incluent l'ensemble des modèles prédictifs infrapersonnels et autonomes ou automatiques qui conditionnent le fonctionnement neurocognitif d'un individu.

la sélection naturelle aurait doté les êtres humains d'un prior adaptatif qui favoriserait *l'alignement de leurs états mentaux avec ceux des autres membres de leur espèce*. Plus précisément, ce *prior d'alignement* correspondrait à une préférence pour des régimes d'inférence active générant des données sensorielles qui indiquent que leurs propres états mentaux sont alignés avec, ou similaires à, ceux des autres membres de leur espèce. Selon eux, la communication coopérative humaine correspondrait à une *stratégie de collecte de données sensorielles* visant à générer des preuves que ce prior adaptatif — la croyance (bayésienne) préalable selon laquelle nos états mentaux sont alignés à ceux des autres — est vrai. Autrement dit, la communication coopérative, qui forme la base d'une grande partie du comportement et de la cognition typiquement humaine, peut-être conçue comme une prophétie auto-réalisante basée sur une prémisse héréditaire (prior adaptatif) selon laquelle nous et les autres membres de notre espèce partageons un phénotype cognitif commun. Ainsi, le phénomène de l'attention jointe serait le résultat de l'implémentation d'un régime d'inférence active héréditaire qui générerait des preuves pour le prior d'alignement, et constituerait la rampe de lancement pour l'émergence de la communication coopérative typiquement humaine, qui consisterait en une extension de ce processus à une échelle temporelle (et informationnelle) plus élevée de l'interaction entre les individus.

Vasil *et al.* (2020) poursuivent avec la constatation selon laquelle la communication coopérative humaine se manifeste typiquement comme un flux d'information circulaire ou bidirectionnelle. Ils affirment que cela est particulièrement vrai pour le type de communication qui s'installe entre un individu et ses donneurs de soin primaires durant l'ontogénie. En continuité avec le paradigme des affordances culturelles (Ramstead *et al.*, 2016), ce flux informationnel serait mis en place par des régimes d'attention partagée — implémentés par le mécanisme d'allocation de la précision — qui guideraient les patterns d'inférence active déployés par les protagonistes. Selon Vasil *et al.* (2020) un élément central du prior d'alignement consiste en la prémisse selon laquelle les actions des agents mènent à une mise à jour des états mentaux des autres

agents. Puisque les états mentaux causent les actions (et par extensions des observations), cela implique que pour collecter des preuves du prior d’alignement, les agents doivent orienter leur attention vers les individus envers lesquels leurs actions sont dirigées. Autrement dit, pour vérifier que nos actions ont bien modifié les états mentaux des autres, nous devons observer les conséquences sensorielles de leurs actions. Les exemples de régimes d’inférence active qui dirigent l’action envers les autres afin de prédire (ou causer) leurs états mentaux incluent les actes gestuels de « pointage » et l’utilisation du langage. Sur la base de travaux récents (Feldman, 2015 ; Hasson *et al.*, 2012 ; Hasson et Frith, 2016), Vasil *et al.* (2020) proposent que lorsque plusieurs individus implémentent de tels régimes d’inférence active afin d’aligner mutuellement leurs états mentaux respectifs, cela donne lieu à l’émergence de *cycles d’action-perception*, décrit comme des boucles d’action-perception intentionnellement *co-construites* qui installent un couplage structurel robuste entre les participants.

Vasil *et al.* (2020) poursuivent en faisant l’hypothèse suivante : au sein de ces cycles d’action-perception qui caractériseraient la communication coopérative humaine, la précision relative accordée aux priors entretenus par chaque agent devrait déterminer les dynamiques de la directionnalité du flux d’information qui s’installe lorsque des individus communiquent. En effet, la précision détermine si les erreurs de prédictions vont être résolues par l’inférence perceptuelle (modification des croyances pour correspondre au monde) ou par l’inférence active (modification du monde pour correspondre aux croyances). Cela implique que lors des échanges communicationnels, les individus possédant des croyances imprécises (incertaines) vont avoir tendance à modifier leur propre modèle du monde pour correspondre à ceux de partenaires qui, par opposition, possèderaient des croyances plus précises (certaines). Sur la base de travaux récents (Karmali *et al.*, 2017 ; Wolpe *et al.*, 2016), Vasil *et al.* (2020) proposent qu’une telle asymétrie dans le flux d’information pourrait capturer les dynamiques des cycles d’action-perception qui caractérisent les interactions entre les enfants humains et leurs donneurs de soins primaires durant le développement.

À partir de propositions exposées précédemment (Fotopoulou et Tsakiris, 2017 ; Friston et Frith, 2015), ils formulent l'hypothèse selon laquelle le couplage récurrent qui s'opère entre enfants et ses donneurs de soins (ainsi que ses pairs plus expérimentés) devrait mener les croyances des individus inexpérimentés à converger vers les causes distales qui génèrent les conséquences sensorielles qu'ils observent, à savoir, les états mentaux de leurs donneurs de soins. Cela les mène à formuler un principe selon lequel les cycles d'actions-perception couplés qui émergent dans de telles relations interpersonnelles seraient caractérisés par un *entraînement asymétrique des priors*. Pour notre part, nous faisons l'hypothèse que ce principe s'applique aussi dans le cas des cycles d'action-perception couplés mis en place qui s'installent dans le cadre de la relation thérapeutique entre un patient et son psychothérapeute.

Vasil *et al.* (2020) poursuivent leur raisonnement en affirmant que l'installation de ces cycles d'action-perception, combinée à la dynamique d'entraînement des priors qui les caractérisent, devrait mener à une synchronisation généralisée des croyances des individus avec celles de leur niche — constituée des états mentaux des autres individus — de telle manière que le modèle génératif incarné par le cerveau des individus en vienne à récapituler la structure et les dynamiques de la niche dans laquelle ils sont imbriqués (c.-à-d.. le modèle génératif moyen incarné le cerveau de leurs congénères). Ce phénomène de synchronisation est le même que celui décrit par Friston et Frith, (2015), et mènerait l'ensemble des individus à converger vers la constitution d'un modèle commun du phénotype et de la niche écologique qu'ils partagent, ce qui inclue les dynamiques de l'échelle cognitive de leur organisation. En effet, le modèle développé par Vasil *et al.* (2020) repose sur le développement d'une conception du domaine informationnel formé par l'ensemble des états mentaux entretenus par une population humaine donnée comme incarnant une échelle à part entière de l'organisation de leur niche écologique partagée. Selon cette perspective, le processus récursif et incrémentiel de modification et de génération des modèles prédictifs incarnés par les cerveaux individuels à travers leur imbrication dans les cycles d'action-

perception couplés qui caractérisent la communication coopérative humaine correspond au mécanisme qui sous-tendrait le phénomène de l'évolution culturelle. Le modèle génératif commun qui émerge de ce processus correspondrait tout simplement à ce que nous appelons *la culture*.

À partir du modèle présenté ici, nous souhaitons formuler l'hypothèse selon laquelle le processus de synchronisation neurocognitive qui s'opère à travers l'installation des cycles d'action-perception couplés qui caractérisent la communication humaine constitue le mécanisme par le biais duquel la psychothérapie psychodynamique permettrait l'apprentissage de la capacité de mentalisation, ainsi que la modification des concepts neurocognitifs qui régissent le fonctionnement émotionnel d'un individu (incluant le concept de soi). Cette hypothèse est basée sur le travail de Gendron et Barrett (2018), qui combinent le modèle des émotions de Barrett (2016) avec le modèle de la communication avancé par Friston et Frith (2015) afin de proposer un modèle du développement collaboratif des concepts émotionnels à travers la synchronisation des dynamiques d'organisation neurocognitive qui s'opère au sein des échanges communicationnels entre deux individus (ou plus). Leur position repose sur la prémisse selon laquelle l'apprentissage émotionnel, ainsi que la majeure partie des épisodes émotionnels vécus par un agent cognitif humain, se déroulent dans le cadre des relations interpersonnelles. Dans ce contexte, ils suggèrent que la perception des émotions devrait comprise comme le processus à travers lequel deux cerveaux (ou plus) en viennent à un accord (ou pas) sur le sens des signaux sensoriels par le biais d'une activité collaborative de conceptualisation de leurs états émotionnels respectifs. En se basant sur le modèle de Friston et Frith (2015), ils précisent leur proposition en affirmant que ce processus pourrait être sous-tendu par une synchronisation dynamique de l'activité de conceptualisation neurocognitive réalisée par deux systèmes prédictifs engagés dans un échange communicationnel.

Selon Gendron et Barrett, le langage jouerait un rôle central dans l'émergence de ce type de synchronie conceptuelle entre les individus. En général, le langage joue un rôle critique dans le domaine des émotions. Durant le développement, les donneurs des soins utilisent (souvent de manière non intentionnelle) des mots pour faciliter l'apprentissage des états mentaux, incluant les concepts émotionnels (Ruffman *et al.*, 2012). À partir de là, les propriétés représentationnelles du langage sont utilisées tout au long de la vie pour organiser la perception des émotions, de telle manière que les disruptions des habiletés langagières puissent mener à des déficits dans ce domaine (Barrett *et al.*, 2016 ; Lindquist *et al.*, 2015). Gendron et Barrett (2018) affirment que les mots servent d'outils mnésiques efficaces afin de permettre la réactivation d'expériences passées (ou de connaissances conceptuelles) pertinentes à l'explication d'un état émotionnel ou d'une situation particulière. Cela est en continuité directe avec la position de Freud sur l'importance du langage, et l'hypothèse de la Neuropsychanalyse (Solms, 2018) ce dernier joue principalement un rôle de support mnésique pour le rappel des souvenirs sous une forme déclarative afin qu'ils puissent être maintenus dans la mémoire de travail. De plus, les mots constituent des outils communicationnels précieux en ce qu'ils permettent aux individus de créer des ponts entre leurs expériences subjectives respectives.

Autrement dit, les mots jouent un rôle critique dans le processus de modélisation prédictive incarné par le cerveau. Dès la naissance, les mots servent à pointer vers les instances dans le monde qui ont une pertinence pour l'allostase, servant d'invitation sociale pour former des concepts (Waxman et Gelman, 2010). Gendron et Barrett (2018) affirment que les mots continuent de jouer un rôle critique tout au long de la vie, car, d'après Lupyan et Clark (2015, p.283) : « *ils sont des sources de priors très flexibles, et peu coûteuses du point de vue métabolique, pour l'architecture prédictive du cerveau* ». Ils poursuivent en affirmant que les mots permettent de créer une cohésion interpersonnelle en facilitant la direction (jointe) de l'attention, en sous-tendant la communication d'intentions, et en organisant l'expérience partagée (Chen et Waxman,

2013 ; Ferry *et al.*, 2010 ; Waxman et Gelman, 2009 ; Waxman et Markow, 1995). De plus, dans le contexte des émotions, l'utilité du langage réside en grande partie en ce qu'il peut servir à résoudre l'ambiguïté (l'incertitude) à propos des états mentaux d'autrui.

En effet, d'après Schilbach *et al.* (2013), pour se bâtir une représentation adéquate des états mentaux d'autrui, la meilleure stratégie consisterait à générer et à tester (à l'aide du langage) des prédictions dans le contexte de l'interaction avec cette personne. Par exemple, un membre d'une dyade (ou d'un groupe) peut tester son modèle de l'état mental d'une autre personne en tentant de nommer l'émotion que cette dernière est entrain de vivre, ce qui revient à verbaliser sa prédiction. De manière très importante, ce type de test pourrait non seulement permettre de raffiner le modèle de l'individu qui teste ses prédictions (inférence perceptuelle), mais aussi changer l'état de son partenaire en modifiant les prédictions de ce dernier quant à son propre état mental (inférence active). Selon le modèle de Vasil *et al.* (2020), chacune de ces alternatives correspond à une preuve pour le prior d'alignement, et c'est la précision relative associée aux modèles respectifs des protagonistes qui détermine laquelle des deux a le plus de chance de se produire (bien qu'il ne s'agisse pas d'un choix dichotomique).

Le type d'interactions langagières décrit ci-dessus pousse Gendron et Barrett (2018) à suggérer que le modèle de synchronie conceptuelle qu'ils avancent serait peut-être mieux décrit comme un modèle de *co-construction conceptuelle*. En effet, ils affirment que la notion de synchronie implique une voie subtile et indirecte vers la perception de l'émotion, basée sur les dynamiques des variations de l'action et des changements intéroceptifs qui ont lieu dans l'interaction avec un partenaire. Selon eux, le terme de co-construction pointe vers un phénomène plus dramatique et direct, de telle manière que les prédictions verbalisées par un individu puissent générer immédiatement un tout nouvel ensemble de prédictions chez leur partenaire. Selon nous, les phénomènes capturés par l'interprétation que font Gendron et Barrett des termes de synchronie et

co-construction coexistent, et correspond à deux formes distinctes que peut prendre le processus de co-conceptualisation émotionnelle qui peuvent s'opérer entre des individus dans le cadre des relations interpersonnelles. Comme nous allons le voir dans la section suivante, nous pensons que l'implémentation de ces deux processus repose sur l'utilisation de deux types différents de mentalisation qui ont été décrits dans la littérature (Fonagy et Luyten, 2012 ; Luyten et Fonagy, 2015) : la mentalisation implicite ou automatique d'une part, et la mentalisation explicite ou contrôlée, d'autre part. Avant d'explorer cette distinction, résumons brièvement le travail réalisé dans cette section.

Le modèle de la communication présenté dans cette section nous offre une explication des mécanismes à travers lesquels les interactions sociales, et plus particulièrement les formes de communication typiquement humaines, en instaurant une pratique collaborative de conceptualisation métacognitive (mentalisation), mènent à une synchronisation interpersonnelle des modèles prédictifs générés par cette dernière, incluant les concepts neurocognitifs du soi et des émotions. À travers l'instauration de cycles d'action-perception couplés entre deux individus (ou plus), la communication coopérative implémente un régime de collecte de preuves sensorielles pour le prior adaptatif selon lequel les états mentaux d'un individu et de ses congénères sont alignés ou similaires (prior d'alignement). Le couplage structurel robuste implémenté par la communication récurrente entre un individu en développement et ses donneurs de soins (ou tous autres congénères plus expérimentés) va mener à un entraînement asymétrique des priors imprécis (incertains) de ce dernier vers les priors plus précis (certains) de ses congénères. Parmi ces priors se trouvent les concepts émotionnels, qui dans ce contexte peuvent être compris comme des priors empiriques, des modèles prédictifs visant la réalisation de l'allostasie qui sont acquis à travers l'expérience, et dont le développement se retrouve donc lui aussi contraint par le prior d'alignement. Selon Vasil *et al.* (2020, p. 10), le prior d'alignement conditionne aussi le développement du concept de soi : « ... within the coupled action-perception cycle of human agents,

evidence for the self amounts to evidence for the other; and evidence for the other is evidence for the self. » D'après Gendron et Barrett (2018), le développement de concepts émotionnels à travers la communication est mieux compris comme un processus de co-conceptualisation au sein duquel le langage jouerait un rôle majeur.

3.2 La psychologie psychodynamique reformulée

3.2.1 L'inconscient dynamique revisité

Jusqu'à présent, nous avons proposé de définir l'espace d'échelle correspondant aux phénomènes en jeu dans la psychothérapie psychodynamique comme un niveau d'auto-organisation particulier des systèmes neurocognitifs humains qui émergerait du couplage structurel avec leur éconiche socioculturelle. Plus précisément, nous avons proposé que le processus principal qui définirait ce niveau d'organisation corresponde au développement et à l'exercice d'une habileté de conceptualisation métacognitive, la capacité de mentalisation, permettant à un individu de modéliser les dynamiques de l'échelle cognitive (ou mentale) de l'activité des êtres humains. Pour l'instant, les modèles que nous avons avancé pour expliquer la psychothérapie psychodynamique éludent l'un des aspects centraux du cadre théorique initial présenté dans notre *explanandum* : les processus mentaux inconscients. Avant de pouvoir procéder à la synthèse de notre modèle explicatif de la psychothérapie psychodynamique, une étape nécessaire donc à développer une compréhension de la spécificité des processus cognitifs inconscients qui seraient en jeu dans l'organisation de l'espace d'échelle défini précédemment.

Conformément aux hypothèses résumées ci-dessus, nous allons nous concentrer sur l'exploration des formes conscientes et inconscientes que peuvent prendre les processus cognitifs qui sous-tendent la capacité de mentalisation. Afin de répondre à certaines critiques considérant le concept de mentalisation comme étant trop large et complexe, Fonagy et Luyten (2012) ont proposé de l'opérationnaliser en divisant ses différents aspects selon quatre dimensions conceptualisées comme des axes polaires : la mentalisation automatique versus contrôlée ; la mentalisation cognitive versus affective ; la mentalisation interne versus externe ; et la mentalisation de soi versus d'autrui. La distinction qui nous intéresse ici est celle qui opposerait une forme de mentalisation automatique ou implicite à une forme de mentalisation contrôlée ou explicite, qui ont été associées à deux ensembles de régions neuronales distincts (Luyten et Fonagy, 2015), qui correspondent en grande partie aux nœuds du réseau du mode par défaut (DMN).

La mentalisation implicite ou automatique serait typiquement inconsciente, non verbale et non réflexive, et correspondrait au type de mentalisation « par défaut » qui sous-tendrait la majorité de nos interactions sociales (Fonagy et Luyten, 2012). Elle serait principalement sous-tendue par l'activité du cortex préfrontal ventromédian, du cortex temporal latéral, et du cortex cingulaire antérieur dorsal, dont l'activation conjointe serait principalement impliquée dans un traitement rapide et automatique de l'information sociale et à la détection de menaces (Satpute et Lieberman, 2006). La recherche empirique suggère que la mentalisation automatique serait déjà acquise autour de la première année de vie (Baillargeon *et al.*, 2010 ; Kovács *et al.*, 2010), ce qui pousse Luyten et Fonagy (2015) à faire l'hypothèse qu'elle serait sous-tendue par des prédispositions innées. La mentalisation explicite ou contrôlée serait quant à elle typiquement consciente, verbale et réflexive et correspondrait à un processus cognitif sériel relativement lent qui demande des efforts conséquents et présente des coûts métaboliques importants (Fonagy et Luyten, 2012). Elle serait principalement sous-tendue par l'activité du cortex préfrontal latéral et médian, du cortex pariétal latéral et

médian, du lobe temporal médian, et du cortex cingulaire antérieur rostral (Lieberman, 2007 ; Satpute et Lieberman, 2006 ; Uddin *et al.*, 2007). La mentalisation explicite apparaîtrait beaucoup plus tard dans le développement (entre 4 et 8 ans ; Carpendale et Lewis, 2006 ; Gweon *et al.*, 2012), et ce parce qu'elle dépendrait du développement des habiletés langagières (Beeghly et Cicchetti, 1994) et des capacités de contrôle exécutif volontaire (Fonagy et Luyten, 2009). Ce faisant, son acquisition serait davantage dépendante du parcours ontogénétique de l'individu (Luyten et Fonagy, 2015).

Cette stratification de la mentalisation en une dimension automatique et inconsciente d'une part, et une dimension contrôlée et consciente d'autre part, n'est pas sans rappeler le célèbre modèle de Kahneman (2013), qui théorise une dichotomie entre deux systèmes de pensée : le système 1, qui sous-tendrait un mode de pensée rapide, instinctif et émotionnel ; et le système 2, qui sous-tendrait un mode de pensée plus lent, plus délibéré, et plus logique. Bien que la dichotomie entre la raison et les émotions soit entièrement incompatible avec le modèle des émotions que nous endossons (Barrett, 2016), la recherche montre que les deux systèmes cérébraux qui sous-tendent la mentalisation implicite et la mentalisation explicite sont impactés différemment par l'intensité affective des expériences émotionnelles dans lesquelles ils sont impliqués. En effet, alors qu'un stress ou une excitation élevés faciliteraient la mentalisation automatique, ils mèneraient à une inhibition des systèmes neuronaux associés avec la mentalisation contrôlée (Lieberman, 2007). Malgré cela, Satpute et Lieberman (2006) suggèrent qu'il existerait une certaine continuité entre ces deux systèmes, et que des apprentissages réalisés à l'aide de la mentalisation explicite, comme des représentations de soi par exemple, pourraient être utilisés par la mentalisation implicite pour diriger la cognition et le comportement. Ils affirment que, comme beaucoup d'autres habiletés cognitives de haut niveau, la mentalisation de scénarios sociaux ou d'états émotionnels particuliers va initialement demander un effort conséquent, mais peut et va s'automatiser progressivement avec la répétition des

expériences. Inversement, nous pensons qu'il est raisonnable faire l'hypothèse selon laquelle un concept émotionnel issu de la mentalisation implicite puisse être récupéré et retravaillé par la mentalisation explicite, à l'aide d'une activité de prise de conscience réflexive et de description verbale des manifestations comportementales et affectives de l'expérience émotionnelle en question.

Précédemment, nous avons proposé que, dans le cadre des approches prédictives du cerveau humain, la capacité de mentalisation puisse être conçue comme d'une activité de conceptualisation métacognitive. Plus précisément, cela implique que la mentalisation correspondrait un processus de génération et d'implémentation de concepts neurocognitifs, des modèles prédictifs qui génèrent des prédictions (perceptuelles et sensorimotrices) afin de minimiser l'erreur de prédiction associée à l'activité affective, comportementale et cognitive encodée par les aires intermédiaires de la hiérarchie cérébrale. Nous pensons que ce modèle peut-être appliqué afin d'explicitier les caractéristiques de la mentalisation automatique et de la mentalisation contrôlée dans le cadre des approches prédictives. Tout d'abord, notons que les deux types de mentalisation instancient des modèles prédictifs relativement complexes qui prédisent les comportements humains en les concevant comme déterminés par des causes distales relativement abstraites : des états mentaux. Cela implique que la mentalisation automatique ne puisse pas être identifiée directement aux processus primaires tels que définis par Carhart-Harris et Friston (2010), qui correspondraient à une forme de cognition désorganisée reliée à une activité limbique et paralimbique débridée. En effet, réalisant une fonction de cognition sociale relativement complexe, et impliquant l'activité de régions corticales reliées aux fonctions exécutives (ex. cortex préfrontal ventromédian), la mentalisation automatique est mieux classée comme faisant partie des processus secondaires. Autrement dit, il est fort probable que son implémentation requiert un degré minimal de régulation de la part des échelles supérieures de la hiérarchie prédictive du cerveau. Selon nous, la différenciation entre ces deux types de mentalisation réside précisément dans le degré ou le niveau de

contrôle exécutif exercé par ces échelles supérieures, qui serait responsable de leurs caractéristiques cognitives et phénoménologiques distinctives.

Comme nous avons pu le remarquer, les aires cérébrales associées à la mentalisation explicite incluent plusieurs régions corticales de haut niveau, comportant non seulement des nœuds du réseau du mode par défaut (ex. cortex préfrontal médian), mais aussi du réseau de contrôle exécutif (cortex préfrontal latéral et cortex pariétal), qui comme nous l'avons vu dans notre *explanans*, est le grand réseau associé aux fonctions exécutives et à la cognition extéroceptive. Ce fait nous pousse à formuler l'hypothèse selon laquelle le caractère contrôlé de la mentalisation explicite pourrait être sous-tendu par une activité collaborative du DMN et du CEN¹². Dans le cadre des approches prédictives, ce contrôle exécutif est conçu comme étant sous-tendu par un investissement soutenu des ressources attentionnelles, implémentant un processus de modulation de la précision accordée aux prédictions descendantes générées par les échelles supérieures de la hiérarchie cérébrale (ex. nœuds du DMN) ou aux erreurs de prédictions ascendantes provenant de l'activité des échelles inférieures et intermédiaires (ex. système limbique et paralimbique, aires sensorimotrices, etc.). D'après Solms et Friston (2018), l'activité de modulation de la précision implémentée par les aires préfrontales serait dirigée par les messages ascendants véhiculés par l'ARAS (et son extension dans le réseau de la saillance), les affects, qui encodent l'incertitude ressentie par l'organisme à propos de ses états internes, et constituent la base de l'expérience consciente. Selon cette perspective, la conscience réflexive de la mentalisation contrôlée résulterait de l'investissement de la conscience affective et des ressources attentionnelles sur le processus de conceptualisation métacognitive afin de le réguler à travers la mise en place d'un contrôle exécutif. Étant donné les demandes computationnelles et les coûts métaboliques élevés d'une telle activité, il n'est pas surprenant qu'elle soit caractérisée sur le plan phénoménologique par l'impression d'être en train de fournir un effort conséquent.

Cependant, les particularités cognitives de la mentalisation explicite ne s'arrêtent pas à son caractère contrôlé et conscient. Un autre aspect central de cette dernière correspond au fait qu'elle est caractérisée par une forme de cognition représentationnelle, et en particulier par l'utilisation du langage. Tel que nous avons pu le voir dans la section précédente, le langage joue un rôle important dans la perception et la régulation des émotions. Par exemple, des études récentes ont montré que de simplement nommer des stimuli visuels émotionnellement chargés avec des appellations symboliques mènerait à une réduction de l'activation de l'amygdale précédemment associée avec ces stimuli, et que cette réduction était associée à l'activation du cortex préfrontal médian (Lieberman, 2007 ; Tabibnia *et al.*, 2008). Ces résultats concordent avec le rôle de suppression de l'activité limbique que Carhart-Harris et Friston (2010) associent au réseau du mode par défaut et au cortex préfrontal médian, et suggèrent que cette activité pourrait être sous-tendue par l'utilisation du langage pour catégoriser (mentaliser) les états émotionnels. Comme proposé par Gendron et Barrett (2018) les mots peuvent être considérés comme des outils représentationnels efficaces afin de permettre la réactivation de priors pertinents à la conceptualisation d'un état émotionnel donné. Selon Solms (2013), les mots serviraient de support représentationnel dont la forme déclarative permettrait le rappel volontaire des informations stockées dans la mémoire, et leur maintien dans la mémoire de travail afin qu'elles puissent être modifiées à l'aide d'un processus de reconsolidation neuronale. Cela est en concordance avec l'association consistante de l'activité du réseau du mode par défaut avec la mémoire autobiographique (Hoffmann, 2020 ; Spreng *et al.*, 2008), qui impliquerait à la fois le rappel de souvenirs épisodiques et leur organisation à l'aide du langage.

Cette remarque nous amène à noter que la mentalisation explicite n'implique fort probablement pas uniquement des mots, mais aussi d'autres formes de représentations extéroceptives pertinentes à la conceptualisation d'un état mental donné, comme des scènes visuelles ou des schèmes proprioceptifs et sensorimoteurs (puisque les concepts

mentaux ou émotionnels impliquent aussi des inférences actives). La nécessité de maintenir et de manipuler ces informations extéroceptives dans la mémoire de travail, et/ou d'implémenter des prédictions sensorimotrices (inférences actives) pourrait expliquer en partie l'activation des nœuds du réseau de contrôle exécutif. Ce type d'activité mentale, dans laquelle un agent cognitif associe des souvenirs épisodiques ainsi que des simulations extéroceptives et proprioceptives à l'utilisation langage afin de se constituer une représentation dynamique de sa personne à travers le temps (autobiographique) peut selon nous être directement associée à ce qui est communément appelé le *soi narratif* (Farmer et Tsakiris, 2012 ; Fivush et Haden, 2003 ; Gallagher, 2000 ; Schechtman, 2011), et que nous considérons comme une extension du soi psychologique. En continuité avec cela, Solms (2013) affirme que l'une des principales fonctions de l'Ego (ou du soi) correspond à une forme de cognition autoréférentielle utilisant des représentations extéroceptives afin d'élaborer des représentations de l'état homéostatique et émotionnel de l'individu. Selon lui, c'est à partir de l'investissement de la conscience affective dans cette activité de cognition représentationnelle qu'émerge la forme de conscience secondaire que l'on appelle la *conscience réflexive* (Edelman, 1992), et à travers laquelle le sujet de l'expérience consciente se prend lui-même pour objet en se représentant ses propres états affectifs, ses comportements, etc. Selon nous, il est clair que l'activité décrite ici correspond directement au type de conceptualisation implémenté par la mentalisation explicite ou contrôlée.

Cependant, Solms (2013) affirme aussi que les processus qui définissent l'organisation de l'Ego ne sont pas tous du type décrit ci-dessus. Selon lui, une partie importante de ces processus se déroulent de manière inconsciente, et ce sont ces processus mentaux inconscients qu'il identifie à l'inconscient dynamique adressé par la psychothérapie psychodynamique. De plus, il conceptualise ces processus mentaux comme des schémas émotionnels automatisés acquis à travers le parcours ontogénétique de l'individu, et en particulier dans le contexte des relations interpersonnelles avec ses

donneurs de soins primaires. Il affirme que le caractère inconscient et automatique de ces schémas ferait en sorte qu'ils restent relativement inertes et qu'ils puissent facilement devenir inadaptés, et donc pathogènes, à travers le temps. Il poursuit en proposant que la spécificité de la psychothérapie psychodynamique réside en ce qu'elle vise principalement à identifier ces schémas émotionnels automatisés, et à les amener à la conscience afin qu'ils puissent être reconsolidés dans la mémoire de travail (Solms, 2018). Nous faisons l'hypothèse que cette partie inconsciente des processus de l'Ego peut être directement associée à la mentalisation implicite. Dans le cadre des approches prédictives, avons proposé que la notion de schéma émotionnel devrait être remplacée par la notion de concept émotionnel proposée par Barrett (2016). Dans le contexte de notre approche prédictive de la capacité de mentalisation, nous proposons que la mentalisation implicite puisse être redéfinie comme le sous-ensemble des processus de conceptualisation émotionnelle qui se déroulent de manière automatique et inconsciente.

De manière consistante avec l'hypothèse formulée ci-dessus, Vatansever *et al.* (2017) ont proposé que le DMN puisse jouer un rôle important dans un mode de « pilote automatique », un type de cognition où la prise de décision est basée sur des modèles prédictifs préétablis. Nous faisons l'hypothèse que la mentalisation implicite appartient à ce type de cognition, et que son caractère inconscient résulte d'un investissement relativement faible des ressources attentionnelles vers la tâche de conceptualisation émotionnelle, ce qui implique une faible modulation de la précision par les fonctions exécutives. Cette hypothèse est appuyée par la moins grande implication des aires corticales supérieures durant la mentalisation implicite (Luyten et Fonagy, 2015). Cependant, cette hypothèse n'implique pas que la précision accordée aux priors (ex. concepts émotionnels) implémentés par ce type mentalisation ne puisse pas être modifiée. En effet, nous avons vu dans notre explanans que le réseau de la saillance (SN), principalement associé à un traitement rapide de l'information affective, joue un rôle important dans la direction des ressources attentionnelles. Dans le contexte des

approches prédictives, si le SN dirige l'attention, alors il module l'allocation de la précision, et ce que les fonctions exécutives soient impliquées ou non. D'après Solms et Friston (2018), si le SN modifie l'allocation de la précision, c'est parce qu'il véhicule l'incertitude ressentie à propos des états internes à travers l'affect. Cela nous mène à formuler l'idée selon laquelle le type d'apprentissage émotionnel réalisé à travers la mentalisation automatique pourrait correspondre à une modulation de la précision des concepts émotionnels implicites à travers un investissement de la conscience affective qui n'impliquerait pas ou très peu les fonctions exécutives, et ne résulterait par conséquent pas en une expérience de prise de conscience réflexive. Par opposition à la mentalisation contrôlée, la mentalisation automatique est *implicite*, car l'expérience subjective — issue de l'investissement de la conscience affective — qui l'accompagne n'impliquerait pas d'un retour réflexif sur l'activité cognitive de conceptualisation émotionnelle ou mentale dans laquelle l'individu est engagé. Dans le langage conceptuel de la phénoménologie merleau-pontienne, on dirait que la mentalisation automatique se déroule sur le mode de la conscience préreflexive (Merleau-Ponty, 2009).

L'idée formulée ci-dessus est en concordance avec le fait qu'une forte intensité affective facilite la mentalisation implicite, et inhibe la mentalisation explicite (Lieberman, 2007). Tel que nous l'avons vu au chapitre précédent, le SN jouerait un rôle causal décisif dans l'activation alternative du DMN ou du CEN (Goulden *et al.*, 2014 ; Menon et Uddin, 2010). En continuité avec le travail de Vatansever *et al.* (2017), ce fait nous pousse à formuler l'hypothèse selon laquelle la mentalisation implicite devrait être caractérisée par une activité cognitive automatique, et qui serait soutenue par une activation relativement isolée du DMN, par opposition à la coactivation du DMN et du CEN que nous avons associé à la mentalisation contrôlée. Selon nous, cette conception de la mentalisation automatique permet de rendre adéquatement compte du type de cognition initialement associé au DMN, à savoir, au type de pensée autoréférentielle et prospective (mind-wandering) qui émerge lorsque nous sommes

dans un état de repos non contraint. Il s'agit de l'activité cognitive dans laquelle nous nous trouvons la plupart du temps : notre mode par défaut.

Dans le contexte des relations interpersonnelles, le type d'apprentissage émotionnel réalisé à travers la mentalisation automatique serait probablement sous-tendu par un processus de synchronisation conceptuelle tel que décrit par Gendron et Barrett (2018) : une modification subtile et indirecte des concepts émotionnels basée sur les dynamiques des variations de l'action (de soi-même et d'autrui) et des changements intéroceptifs (c.-à-d., affectifs) qui ont lieu dans l'interaction entre les individus. Dans le même ordre d'idée, le type d'apprentissage émotionnel réalisé à travers la mentalisation contrôlée serait probablement sous-tendu par un processus de co-construction conceptuelle tel que décrit par Gendron et Barrett (2018) : une modification plus dramatique et directe des concepts émotionnels opérée par une forme de communication langagière dans laquelle les prédictions verbalisées par un individu peuvent générer immédiatement un tout nouvel ensemble de prédictions chez leur partenaire.

À partir du travail réalisé jusqu'à présent, nous pensons être capables de proposer une reconceptualisation des processus mentaux inconscients qui sont impliqués dans la psychothérapie psychodynamique. Plus précisément, nous pensons être à même d'opérationnaliser le concept psychanalytique *d'inconscient dynamique*, en le délimitant à un sous-ensemble spécifique des processus qui composent l'inconscient cognitif. Pour ce faire, nous commencerons par une brève révision du sens que la théorie psychanalytique attribue à ce concept. Selon la psychanalyse freudienne, l'étiologie des névroses reposerait sur l'isolation de contenus psychiques tels que des motivations pulsionnelles, qui développeraient une autonomie par rapport à la personnalité consciente de l'individu (l'Ego), et qui pourraient influencer l'activité mentale et le comportement de ce dernier indépendamment de son contrôle volontaire. Ce sont l'ensemble de ces groupes psychiques isolés qui forment ce que la théorie

psychanalytique appelle l'inconscient dynamique, compris comme la partie inconsciente des processus mentaux qui déterminent le fonctionnement psychologique d'un individu. Selon cette perspective, ce sont les conflits intrapsychiques entre ces groupes psychiques inconscients et la structure régulatrice de l'Ego qui seraient à l'origine de la souffrance émotionnelle et des symptômes qui caractérisent les formes de psychopathologies adressées par la psychothérapie psychodynamique.

Dans ce contexte, l'Ego est conçu comme une instance psychique ayant pour fonction de réduire et de contenir l'excitation causée par la demande de travail constante que les motivations pulsionnelles exercent sur le système nerveux, ce qu'il peut réaliser de différentes manières. La première consiste intégrer cette excitation et à la traduire en une activité comportementale ou mentale adaptative, en réalisant les actions permettant d'assouvir les besoins biologiques dont elle constitue l'expression, et/ou en utilisant le langage afin d'opérer une prise de conscience réflexive de l'expérience affective à laquelle elle correspond afin de la réguler. Lorsque les motivations pulsionnelles exercent des demandes que l'Ego est incapable de convertir en une activité comportementale ou cognitive adaptative, et/ou que l'intensité affective qui leur est liée est jugée comme étant intolérable, l'Ego serait contraint de réprimer l'expression de ces dernières à l'aide de ce que l'on appelle des mécanismes de défense. Le plus important d'entre eux, le mécanisme de refoulement consisterait en un acte mental qui exclut systématiquement les contenus mentaux problématiques de l'expérience consciente du sujet, ce qui aurait pour effet d'éviter d'éliciter la charge affective qui leur est liée, et de maintenir l'excitation du système nerveux à un niveau acceptable. Selon la psychanalyse freudienne, l'isolation des groupes psychiques autonomes qui vont former l'inconscient dynamique est la conséquence de l'application du refoulement. Dans ce contexte, l'inconscient dynamique est donc le produit d'une activité récurrente d'isolation de contenus mentaux de l'expérience consciente d'un individu.

Comme nous l'avons vu précédemment, Solms a proposé que les groupes psychiques isolés qui composent l'inconscient dynamique soient redéfinis comme des schémas émotionnels inconscients, et que ces derniers puissent être considérés comme une partie implicite et automatisée des processus de l'Ego. Pour notre part, nous avons proposé de reconceptualiser ces schémas émotionnels inconscients comme des concepts émotionnels implicites issus de l'exercice de la capacité de mentalisation automatique. Sur la base de cette proposition, nous formulons l'hypothèse selon laquelle la mentalisation automatique correspond directement aux processus cognitifs inconscients qui sont visés par la psychothérapie psychodynamique. *Plus précisément, nous proposons que l'inconscient dynamique puisse être redéfini comme la partie automatique et implicite des processus cognitifs qui sont en charge de la fonction de conceptualisation métacognitive permettant l'autorégulation émotionnelle et la cognition sociale, incluant les concepts émotionnels et mentaux qu'elle génère et qu'elle implémente.* Performés par l'activité relativement isolée des nœuds du DMN (Luyten et Fonagy, 2015), ces processus cognitifs correspondraient à un mode de « pilote automatique » de la fonction de conceptualisation émotionnelle (Vatansever *et al.*, 2017), qui se manifesterait sur le plan phénoménologique sur le mode de la conscience préreflexive.

Selon nous, les concepts émotionnels instanciés à travers la mentalisation implicite correspondent à des priors empiriques dont une grande partie ont été générés à partir des expériences ontogénétiques précoces qui ont eu lieu dans le cadre des relations interpersonnelles intimes entre un individu et ses donneurs de soins primaires. Dû à l'immaturation neurologique prolongée qui caractérise les enfants humains, ces expériences n'ont bien souvent pas pu être enregistrées dans la mémoire épisodique, et les concepts émotionnels auxquels elles donnent lieu prennent donc majoritairement une forme implicite et procédurale (Solms, 2018). Néanmoins, ces priors empiriques précoces forment la base du fonctionnement mental de l'individu, et leur application récurrente tout au long du développement fait en sorte qu'ils subissent un apprentissage

procédural profond, ce qui résulterait en une forte automatisation et une résistance au changement importante. Il n'est donc pas surprenant qu'ils puissent finir par ne pas résulter en un comportement ou une cognition adaptative dans l'environnement actuel de l'individu. Comme nous l'avons vu, la mentalisation implicite est la forme dominante de mentalisation, et cette domination est accrue lorsque le système prédictif est soumis à un niveau d'intensité affective élevé. Cela implique que le fonctionnement émotionnel d'un individu, et particulièrement dans le contexte de situations à haute teneur affective (ex. relations d'attachement), soit en grande partie déterminé par l'implémentation automatique de concepts émotionnels par la mentalisation implicite.

Comme nous l'avons vu au chapitre précédent, l'intensité affective correspond à l'incertitude ressentie par un organisme quant au maintien de ses états internes dans les normes définies par son phénotype (Solms et Friston, 2018). Selon nous, la surdétermination des conditions du milieu interne et externe de l'organisme devrait mener à une grande variabilité dans l'incertitude ressentie vis-à-vis de situations données. Autrement dit, et même lorsque que des situations sont caractérisées par un état global du système cerveau-corps-environnement objectivement similaire, il est possible que le système prédictif, sur la base de variations minimales des entrées sensorielles qu'il reçoit, les évalue comme présentant un niveau d'incertitude significativement différent. Sur la base de ce constat, nous faisons l'hypothèse qu'un même système prédictif puisse entretenir des modèles prédictifs différents sur des ensembles d'entrées sensorielles très similaires. Comme l'engagement sélectif dans la mentalisation implicite ou explicite dépend en grande partie de l'intensité affective ressentie à un moment donné, les considérations précédentes nous poussent à penser qu'à travers le développement émotionnel d'un individu, ces deux types de mentalisation puissent avoir été alternativement appliqués à des situations similaires, menant à l'élaboration de concepts émotionnels différents portant sur des ensembles d'entrées sensorielles quasi identiques. Dans ce contexte, nous pensons qu'il est possible qu'un ensemble de concepts émotionnels implicites et automatiques puissent

coexister de manière parallèle à des concepts émotionnels explicites et contrôlés portant sur des situations et des domaines sensoriels similaires.

La différence entre ces deux types de concepts émotionnels réside dans les niveaux de l'activité de la hiérarchie prédictive incarnée par le cerveau qu'ils impliquent. Les concepts émotionnels implicites et automatiques impliquent une activité relativement autonome des échelles intermédiaires et inférieures de la hiérarchie prédictive. Les concepts émotionnels explicites et contrôlés impliquent quant à eux l'activité de la hiérarchie prédictive dans son ensemble, orchestrée par le contrôle exécutif fourni par l'implication des échelles supérieures. Ainsi lorsque nous disons que ces concepts coexistent de manière parallèle, nous parlons d'un parallélisme dynamique, impliquant que deux (ou plus) concepts métacognitifs d'échelles différentes portant sur le même domaine d'erreurs de prédiction peuvent cohabiter dans le même système prédictif, mais qu'ils ne peuvent pas être implémentés en même temps, car ils génèrent des prédictions différentes. Autrement dit, ils sont en compétition pour la minimisation de ce domaine d'erreurs de prédiction.

En partant de l'exemple de la rivalité binoculaire, Hopkins (2012) a proposé que dans le cadre des approches prédictives du cerveau humain, la notion de *conflit intrapsychique* puisse être reformulée comme la dynamique de compétition qui peut s'installer lorsque deux modèles prédictifs incompatibles sont entretenus par un même système prédictif. En combinant cette hypothèse avec celles que nous avons formulées dans le paragraphe précédent, nous proposons que la notion de conflit intrapsychique puisse être redéfinie comme une forme de compétition qui s'installerait entre des concepts neurocognitifs différents incarnant des modèles prédictifs qui visent à minimiser un même ensemble d'erreurs de prédiction en implémentant des prédictions différentes. Plus précisément, nous faisons l'hypothèse que dans le contexte de l'autorégulation émotionnelle, cette forme de compétition s'installe majoritairement entre les concepts émotionnels générés et implémentés par les deux types de

mentalisation présentés dans cette section : la mentalisation implicite et la mentalisation contrôlée.

Comme nous le savons, l'activité d'un système prédictif est dirigée par le processus d'allocation de la précision, qui représente l'incertitude associée aux prédictions et aux erreurs de prédiction. Ce faisant, nous proposons que l'engagement sélectif dans la mentalisation automatique ou contrôlée soit déterminé par une évaluation de la précision des concepts émotionnels qu'elles implémentent, et qui correspond à une prédiction de leur capacité relative à minimiser efficacement les erreurs de prédiction auxquelles le système est confronté. Cette proposition nous permet de formuler une hypothèse quant à la raison de la domination générale de la mentalisation implicite sur la mentalisation explicite, ainsi que de l'exagération de cette dominance reliée à un haut niveau d'incertitude ressentie par le système. Dans ce contexte, nous proposons que la mentalisation explicite présente un désavantage important, car il s'agit d'un processus cognitif de haut niveau qui présente des coûts métaboliques et computationnels relativement plus élevés. Comme beaucoup de processus cognitifs de haut niveau, elle implique l'investissement soutenu des ressources attentionnelles, ainsi que la collaboration d'un plus grand nombre de zones cérébrales distinctes (Luyten et Fonagy, 2015). Ce faisant, elle correspond à un processus cognitif sériel qui demande l'investissement global du système prédictif dans la tâche de conceptualisation métacognitive, ce qui implique d'interrompre toute autre activité cognitive ou comportementale complexe. Ces remarques suffisent à elle seule pour justifier la domination générale de la mentalisation implicite, mais un examen plus attentif peut nous permettre de mieux saisir l'inhibition de la mentalisation explicite en lien à une intensité affective élevée.

Comme de nombreuses activités cognitives de haut niveau qui sont le propre de l'espèce humaine, la mentalisation explicite correspond à un processus de traitement d'informations d'un niveau d'abstraction élevé, modélisant des phénomènes se

déroulant à des échelles spatiotemporelles relativement grandes, une activité qui présente un haut niveau d'incertitude (Carhart-Harris et Friston, 2010). De plus, ses demandes computationnelles importantes impliquent la suspension de l'activité cognitive et comportementale reliée à épisode émotionnel sur lequel elle porte, ce qui tel que nous l'avons vu résulterait de l'exercice d'un contrôle exécutif sur l'activité de conceptualisation émotionnelle automatique réalisée par la mentalisation implicite. Cependant, cette activité avait été implémentée pour répondre à l'impératif de minimisation des erreurs de prédiction intéroceptives reçues par le système prédictif, et en attendant que la mentalisation explicite fasse son travail, ce dernier est toujours soumis à une élévation de l'incertitude qu'il ressent par rapport au processus de régulation de ses états internes. Autrement dit, la mentalisation explicite nécessite que le système soit capable de tolérer un niveau d'incertitude relativement élevé — ce qui correspond à tolérer une déviation significative de ses valeurs homéostatiques préférées — le temps de trouver une manière plus efficace de conceptualiser (minimiser) les erreurs de prédiction qui en sont responsable.

En d'autres termes, la mentalisation explicite est une sorte de pari, qui implique de tolérer une augmentation temporaire de l'énergie libre du système, en prédisant que cela va mener à une meilleure minimisation globale de cette dernière à une échelle temporelle plus élevée. Dans le cadre de la théorie des systèmes dynamiques, cette tolérance momentanée d'une augmentation relative de l'entropie interne du système correspond à ce que l'on appelle la stabilité (ou l'instabilité) dynamique, ou *métastabilité*, une stabilité globale qui émerge d'interaction dynamique entre des processus instable (Friston, 2000). Dans le cas d'un système vivant, la métastabilité serait cruciale, car elle lui permettrait de s'engager dans une dynamique d'activité exploratoire lui offrant la possibilité de se constituer un répertoire flexible d'états globaux, et d'éviter de se retrouver dans un ensemble d'états trop limité ne lui permettant pas de s'adapter efficacement à son environnement (Thompson, 2007). Dans ce contexte, nous pourrions dire que la métastabilité permet *l'apprentissage à*

travers l'expérience (Bion, 1984 ; Solms et Friston, 2018), ce qui correspond à une modification graduelle de l'organisation du système nécessaire au maintien d'un couplage structurel robuste avec sa niche écologique, et ce malgré les caractéristiques changeantes de cette dernière. Dans le cadre de la mentalisation explicite, cela implique de moduler l'activité de conceptualisation émotionnelle automatique afin de générer de nouvelles prédictions permettant de minimiser plus efficacement les erreurs de prédiction qu'elle tente d'expliquer, ce qui correspond à un processus exploratoire d'implémentation et d'évaluation de nouvelles inférences actives et perceptuelles.

Nous faisons l'hypothèse selon laquelle, lors de la survenance d'expériences affectives intenses, le système ne peut pas s'engager dans le type d'activité exploratoire implémenté par la mentalisation explicite, car le niveau élevé d'incertitude ressentie est jugé comme étant intolérable et demande à être réduit au plus vite. Dans ce contexte, il est logique que le système ait tendance à favoriser l'engagement sélectif dans une activité de mentalisation implicite, permettant la mise en place immédiate de mesures visant à réduire les erreurs de prédiction à un niveau acceptable. Selon nous, si ces expériences affectives se répètent, leur conceptualisation à travers l'exercice récurrent de la mentalisation implicite devrait avoir deux conséquences majeures. La première serait de causer un renforcement de la précision des prédictions implémentées par la mentalisation implicite, qui bien que pouvant être relativement imparfaites, correspondent à la meilleure solution que le système a trouvée pour réduire l'incertitude ressentie à un niveau tolérable. La deuxième serait d'inhiber systématiquement la mentalisation explicite de ces expériences affectives, ce qui aurait pour effet de maintenir les prédictions qu'elle pourrait potentiellement générer à un niveau de précision relativement bas. Prises ensemble, ces deux conséquences auraient pour effet d'accentuer la dominance de la mentalisation implicite, en faisant en sorte que, lors la survenance de certaines expériences affectives intenses, ses prédictions soient systématiquement privilégiées par rapport à celles qui pourraient être générées par la

mentalisation explicite, et ce car elles présenteraient presque toujours un meilleur niveau de précision.

D'après le travail réalisé jusqu'à maintenant, nous proposons que malgré le développement d'une capacité de mentalisation explicite convenable, certains concepts émotionnels implicites reliés à des expériences présentant une intensité affective importante puissent rester relativement hermétiques à une modulation à travers l'exercice d'un contrôle exécutif. Néanmoins, il est possible que dans des situations qui, bien qu'étant objectivement similaires, n'ont pas élicité une élévation de l'incertitude ressentie trop importante, la capacité de mentalisation contrôlée ait pu être exercée pour développer des concepts émotionnels explicites portant sur la minimisation d'ensembles d'erreurs de prédictions sensiblement équivalents à ceux sur lesquels portent les concepts émotionnels implicites mentionnés ci-dessus. C'est dans le contexte de la survenue de ces ensembles d'erreurs de prédictions particuliers que devrait selon nous émerger le phénomène de conflit intrapsychique, au sein duquel les deux échelles de la capacité de mentalisation vont entrer en compétition pour le contrôle de l'activité de conceptualisation métacognitive qui gouverne l'autorégulation émotionnelle et la cognition sociale. Étant donné le renforcement qu'ont probablement subi les concepts émotionnels implicites à travers la répétition de réactions affectives intenses à ces situations, ils vont posséder une précision plus élevée que les concepts émotionnels explicites, et vont être privilégiés dans la plupart des cas.

Étant donné le caractère relativement lent du type d'apprentissage émotionnel que nous avons associé avec la mentalisation implicite²¹, nous faisons l'hypothèse que la

²¹ En effet, alors que la mentalisation implicite est performée très rapidement, le type d'apprentissage émotionnel qu'elle sous-tend reposerait selon nous sur les mécanismes décrits par Gendron et Barrett (2018) : une modification subtile et indirecte des concepts émotionnels basée sur les dynamiques des

domination de cette dernière pourrait mener à une stagnation des concepts émotionnels reliés à des situations ayant une tendance à générer un haut degré d'incertitude ressentie, ce qui aurait pour effet de limiter le fonctionnement émotionnel d'un individu à un ensemble d'états sous-optimaux pouvant mener à des comportements inadaptés. L'inefficacité relative de ces comportements — suffisamment efficaces pour réduire l'incertitude ressentie à un niveau tolérable, mais pas assez pour fournir une solution durable aux problèmes qui ont mené à son augmentation excessive — pourrait mener à une boucle de rétroaction au sein de laquelle l'intensité affective reliée à ces situations serait toujours maintenue à un niveau trop élevé pour permettre l'engagement dans la mentalisation explicite, renforçant ainsi la domination de la mentalisation implicite, et ainsi de suite.

Dans le contexte de notre reformulation de la notion de conflit intrapsychique, nous pensons pouvoir identifier ce que Freud appelle la *personnalité consciente* de l'individu à l'ensemble des concepts neurocognitifs issus de l'activité de conceptualisation métacognitive explicite correspondant à l'exercice de la mentalisation contrôlée. Nous faisons l'hypothèse que le complexe, formé par l'intégration et la collaboration dynamique de ces concepts correspond directement à ce que l'on appelle communément le *Soi psychologique*. Les concepts métacognitifs développés à travers l'exercice de cette capacité formeraient la base de la compréhension qu'un individu se fait de son fonctionnement émotionnel et psychologique sur le mode de la conscience réflexive. Ces derniers formeraient une partie importante de sa vie mentale, pouvant être volontairement implémentés à travers la manipulation des représentations langagières et des souvenirs épisodiques qui leur sont liés. Cependant, lorsque sujet à

covariations de l'action (de soi-même et d'autrui) et des changements intéroceptifs (c.-à-d., affectifs) qui ont lieu dans l'interaction d'un individu et sa niche écologique, incluant son environnement social.

des réactions affectives intenses en lien avec certaines situations problématiques, l'individu se trouverait momentanément incapable d'exercer sa capacité de mentalisation explicite, s'engageant involontairement dans l'implémentation automatique de concepts émotionnels sur lesquels il est incapable d'exercer un contrôle exécutif. Dans le cas des concepts émotionnels précoces développés dans ses premières années de vie, l'individu serait dépourvu de souvenirs épisodiques lui permettant d'attacher des représentations déclaratives sur les prédictions qu'ils implémentent, ce qui renforcerait leur opacité et leur herméticité à une modulation potentielle par les fonctions exécutives.

Nous faisons l'hypothèse selon laquelle l'alternance dynamique entre l'implémentation de modèles prédictifs différents portant sur les mêmes ensembles d'erreurs de prédictions, causée par la variabilité de l'incertitude ressentie vis-à-vis d'instances différentes de situations similaires, devrait résulter en une augmentation générale de l'énergie libre du système prédictif due à l'imprédictibilité, et donc l'incertitude, du fonctionnement émotionnel de l'individu. Cela associé au potentiel maladaptatif de l'inertie relative des concepts émotionnels implicites correspond selon nous à la cause principale du phénomène de détresse psychologique que la psychologie psychodynamique capture à travers le concept de « souffrance émotionnelle ». Nous proposons que ce phénomène corresponde à l'expérience d'un haut degré d'incertitude quant au processus de réalisation de l'allostase qui résulte d'un dysfonctionnement des habiletés de conceptualisation métacognitives qui sous-tendent les capacités d'autorégulation émotionnelles. Ce dysfonctionnement serait caractérisé par la dominance de modèles prédictifs favorisant l'évitement proximal de l'augmentation temporaire de l'incertitude ressentie au détriment de l'installation d'une dynamique de métastabilité qui permettrait une meilleure minimisation de l'énergie libre globale du système à des échelles temporelles plus élevées. En concordance avec notre conception de la souffrance émotionnelle, Joffily et Coricelli (2013) ont avancé un modèle de la valence émotionnelle selon le FEP dans lequel ils directement la valence émotionnelle

négative avec une augmentation de l'énergie libre d'un système prédictif. Nous pensons que le type d'augmentation globale et temporellement étendue de l'énergie libre que nous associons au caractère suboptimal des modèles prédictifs implémentés par la mentalisation automatique pourrait expliquer la valence émotionnelle négative qui caractérise les troubles de l'humeur, comme les troubles dépressifs ou bipolaires.

Selon nous, cette reformulation de la souffrance émotionnelle peut être directement utilisée pour reconceptualiser et opérationnaliser le concept de refoulement dans le cadre des approches prédictives du cerveau humain. D'après Solms (2013, 2018), le refoulement correspond à une automatisation prématurée de schémas émotionnels inefficaces qui ne permettent pas de remplir efficacement les besoins de l'individu dans son interaction avec son environnement. Nous pensons que cette formulation est problématique, car elle suggère une perspective *top-down* similaire à celle proposée par Satpute et Lieberman (2006), qui consiste à proposer que les apprentissages émotionnels s'opèreraient dans un premier temps par le biais d'une activité cognitive contrôlée, et que la répétition ferait en sorte qu'ils seraient ensuite consolidés à travers le temps jusqu'à être automatisés. Bien que nous soyons ouverts à la possibilité de l'automatisation de certains concepts émotionnels développés originellement à travers la mentalisation explicite, nous pensons cependant que les faits présentés jusqu'à présent suggèrent qu'une approche opposée, ou *bottom-up*, serait plus adéquate.

En effet, la recherche suggère que ce soit la mentalisation implicite qui gouverne la majorité de notre fonctionnement émotionnel (Fonagy *et al.*, 2018 ; Fonagy et Luyten, 2012 ; Luyten et Fonagy, 2015). Ce constat nous mène formuler l'hypothèse selon laquelle le mode d'apprentissage émotionnel « par défaut » correspond à un processus de conceptualisation neurocognitive automatique et implicite, et non pas au type de cognition explicite et contrôlée telle décrit par Satpute et Lieberman (2006). Cela nous mène à penser que la majorité des modèles prédictifs qui gouvernent le fonctionnement émotionnel d'un individu correspondent à des concepts émotionnels implicites.

Comme nous l'avons proposé précédemment, la mentalisation explicite correspondrait à l'exercice d'une forme de cognition représentationnelle permettant de générer des prédictions *top-down* implémentant un contrôle exécutif sur l'activité de conceptualisation émotionnelle automatique. Ce faisant, nous faisons l'hypothèse que la majorité des concepts émotionnels explicites implémentés par la mentalisation contrôlée seraient issus de la récupération et de la modulation de concepts émotionnels implicites ayant été générés par la capacité de mentalisation automatique. Dans ce contexte, nous proposons que le phénomène du refoulement ne corresponde pas à une automatisation précoce des concepts émotionnels, mais plutôt à une tendance du système prédictif à favoriser l'implémentation de concepts émotionnels qui sont d'ores et déjà automatiques, et qui ont été générés comme tels à travers l'exercice de la mentalisation implicite.

Dans le cadre du modèle de la souffrance émotionnelle avancé précédemment, nous proposons de reconceptualiser le phénomène du refoulement comme un régime d'allocation de précision favorisant l'implémentation de prédictions qui visent à la minimisation immédiate de l'incertitude ressentie (affect) à un niveau tolérable pour le système prédictif. Plus précisément, dans le cadre du type de pensée autoréférentielle dans laquelle nous nous trouvons la plupart du temps (Raichle *et al.*, 2001), le refoulement correspondrait à un régime de direction des ressources attentionnelles guidant l'activité mentale de l'individu afin d'éviter l'instanciation de représentations ou de concepts qui pourraient entraîner l'émergence d'une expérience présentant une intensité affective trop importante (Eriksen et Kuehe, 1956). De ce fait, le refoulement correspond selon nous à la manifestation ou à la contrepartie cognitive des mêmes mécanismes qui sont en jeu dans le phénomène comportemental bien documenté que l'on appelle *l'évitement*, et qui correspond à la tendance à ne pas s'exposer à des situations environnementales anxigènes ou stressantes (Friedman et Silver, 2007 ; Tiet *et al.*, 2006 ; Zeidner et Endler, 1996). Dans le sens que nous lui attribuons ici, le phénomène du refoulement est une manifestation directe de la manière dont les

systèmes prédictifs que sont les cerveaux humains cherchent à minimiser leur énergie libre à travers un processus d'évitement proximal de la surprise (Friston, 2010 ; Friston et Klaas E., 2007).

Nous pensons que la reconceptualisation des concepts d'inconscient dynamique, de souffrance émotionnelle, et de refoulement proposée dans cette section peut nous permettre de procéder à une reformulation de l'un des objectifs principaux de la psychothérapie psychodynamique. Dans notre *explanandum*, nous avons vu que Freud définissait l'un des objectifs centraux de la cure psychanalytique comme étant la prise de conscience des contenus psychiques inconscients. Solms (2018) a proposé que cet objectif puisse être reformulé comme la désautomatisation des schémas émotionnels inconscients à travers leur prise de conscience réflexive de leurs manifestations comportementales et cognitives. À l'aide des hypothèses formulées dans cette section, nous pouvons redéfinir cet objectif comme *la transition d'un type de mentalisation implicite et automatique vers un type de mentalisation explicite et contrôlé*. Cette transition devrait passer par la direction des ressources attentionnelles sur le processus de conceptualisation métacognitive afin d'y exercer un contrôle exécutif permettant de le réguler et de modifier les concepts émotionnels et mentaux qu'il génère et qu'il implémente. Le processus décrit ci-dessus reposerait sur une activité de communication coopérative pratiquée par un patient et son psychothérapeute, et au sein de laquelle ces derniers procéderaient à une co-construction des concepts métacognitifs utilisés par le patient pour comprendre son fonctionnement cognitif et celui d'autrui.

Dans ce contexte, nous pensons que les deux techniques principales de la psychothérapie psychodynamique, l'interprétation des associations libres et du transfert, peuvent-être aisément reconceptualisées comme ayant pour effet de travailler à remplir cet objectif. En effet, elles correspondent respectivement à une activité collaborative d'observation et de la conceptualisation métacognitive explicite des manifestations cognitives et comportementales du fonctionnement émotionnel de

l'individu. En plus d'entraîner le patient à l'exercice de sa capacité de mentalisation explicite, les expériences vécues dans le cadre de la relation thérapeutique fournissent à ce dernier de nouveaux souvenirs épisodiques qu'il va pouvoir utiliser pour se représenter les processus cognitifs qui sous-tendent l'implémentation automatique de certains de ses concepts émotionnels implicites pathogènes les plus anciens et les plus tenaces. De plus, la recherche suggère que la simple présence d'une personne de confiance permettrait de moduler l'expérience affective et la douleur physique ressentie par un individu (Atzil et Barrett, 2017 ; Fotopoulou et Tsakiris, 2017 ; Krahé *et al.*, 2013 ; Ochsner et Gross, 2008), un effet qui serait potentiellement sous-tendu par l'activité du DMN (Amft *et al.*, 2015 ; Xie *et al.*, 2016). Cela suggère que le support psychologique fourni par le psychothérapeute pourrait aider à réduire l'intensité affective vécue par le patient à un niveau tolérable, et ainsi favoriser l'exercice de la mentalisation explicite.

Dans cette section, nous avons proposé d'identifier l'activité mentale inconsciente adressée par la psychothérapie psychodynamique à un sous-ensemble particulier des processus cognitifs qui composent l'inconscient cognitif, à savoir ceux qui implémentent l'activité de mentalisation implicite et automatique qui sous-tendrait la majeure partie de notre fonctionnement émotionnel et interpersonnel. Sur la base de cette hypothèse, nous avons proposé de redéfinir l'objectif principal de la psychothérapie psychodynamique comme étant d'assister et d'entraîner le patient à l'exercice de sa capacité de mentalisation explicite et contrôlée, dans le but de lui permettre d'apprendre à exercer un contrôle exécutif sur son fonctionnement émotionnel. Afin de pouvoir fournir une explication satisfaisante de la psychothérapie psychodynamique, il nous faut nous constituer un modèle de la manière dont les interactions interpersonnelles telles que celles qui occurred entre un psychothérapeute et son client puissent mener à une amélioration de la capacité de mentalisation de ce dernier. Pour ce faire, nous devons développer une compréhension de la manière dont

cette dernière est acquise à travers les relations interpersonnelles durant le développement.

3.2.2 Le phénomène de l'attachement selon le principe de l'énergie libre et les approches prédictives du cerveau humain

Nous faisons l'hypothèse que parmi les premiers concepts émotionnels implicites générés à partir de l'exercice de la mentalisation automatique dans le cadre des relations interpersonnelles qui se déroulent durant l'ontogénie se trouvent les schémas d'attachement (Ainsworth *et al.*, 1978). Selon la théorie de l'attachement, chaque enfant est génétiquement prédisposé à chercher une relation de dépendance et de support émotionnel avec un donneur de soin primaire (généralement un parent), envers lequel il se tourne pour trouver confort et sécurité face au stress et aux menaces perçues (Bowlby, 1999). Cittern *et al.* (2018) proposent un modèle du phénomène de l'attachement dans le cadre du FEP et des approches prédictives du cerveau humain, et définissent les schémas d'attachement comme des modèles prédictifs de représentation du niveau de confiance accordé par un individu au support potentiel de ses donneurs de soins primaires dans la tâche de régulation de ses états homéostatiques et affectifs. D'après le modèle de la mentalisation incarnée (Fotopoulou et Tsakiris, 2017) présenté précédemment, la réalisation de l'allostasie d'un enfant humain est un processus collaboratif au sein duquel ses donneurs de soins jouent le rôle principal en implémentant les inférences actives permettant de minimiser les erreurs de prédictions intéroceptives qui encodent les variations de ses états homéostatiques. De manière similaire, Cittern *et al.* (2018) partent de la prémisse selon laquelle l'impératif de minimisation de l'énergie libre qui s'impose aux enfants humains les poussent à modéliser — par l'inférence perceptuelle *et* active — les processus causaux qui sont

responsables de la minimisation de leurs erreurs de prédictions intéroceptives, et qui dans les premières années de vie, correspondent majoritairement aux actions de leurs donneurs de soins primaires.

Si l'on suit les hypothèses de la théorie de l'attachement, étant donné l'altruisme et la dépendance prononcée des nouveau-nés humains, il est fort probable que l'enclenchement de cette modélisation soit facilité par un ensemble de priors adaptatifs biologiquement héréditaires, un « système d'attachement », composé notamment de réflexes communicationnels comme le fait de pleurer par exemple, qui peut être conceptualisé dans le contexte des approches prédictives comme une inférence active de communication affective ayant pour but de solliciter les inférences actives de régulation homéostatique des donneurs de soins potentiels. De plus, la théorie de l'attachement stipule que le « système d'attachement » aurait coévolué avec un autre système comportemental chez les donneurs de soins, le « système de *caregiving* », défini de manière générale comme la tendance à protéger, à reconforter, et à prendre soin de ses congénères (George et Solomon, 2008 ; Solomon et George, 1996). Ces deux systèmes assureraient que l'enfant et ses donneurs de soins soient mutuellement prédisposés à rechercher et à maintenir une proximité les uns aux autres (Cittern *et al.*, 2018). Ils auraient été cosélectionnés à cause de leur efficacité pour assurer la survie de l'enfant, ainsi que le maintien de ce dernier à un niveau de stress (d'incertitude, de surprise, etc.) optimal pour favoriser son développement neurocognitif (Cittern *et al.*, 2018).

D'après Cittern *et al.* (2018), le processus collaboratif de minimisation des erreurs de prédiction intéroceptives de l'enfant mènerait à la génération chez ce dernier d'un prior empirique, un schéma d'attachement, qui encoderait les attentes de l'individu à propos du support potentiel de ses donneurs de soins primaires dans la tâche de régulation de ses états internes, ainsi que le type d'inférences actives qu'il doit mettre en place afin de minimiser les erreurs de prédictions liées aux interactions avec ces derniers. Ce

prior devrait par la suite s'appliquer à l'ensemble des relations interpersonnelles intimes entretenues par l'individu tout au long de sa vie. Les études réalisées pour sonder le style d'attachement chez les adultes démontrent une stabilité relativement forte du type d'attachement à travers le développement (Hesse, 2008 ; Main *et al.*, 1985). De plus, il semble que le type d'attachement prénatal d'un parent soit un bon prédicteur du type d'attachement de leur enfant (van IJzendoorn, 1995). Certains travaux ont associé le développement d'un attachement insécuré ou ambivalent à certaines caractéristiques problématiques dans la communication affective des donneurs de soins primaires de l'enfant, qui démontre par exemple des messages incohérents et un décalage avec l'expérience affective de ce dernier (Lyons-Ruth *et al.*, 1999 ; Madigan *et al.*, 2006). D'après Cittern *et al.* (2018), ces travaux suggèrent une tendance vers la transmission intergénérationnelle des types d'attachement (bien que cette dernière ne soit pas inévitable, surtout s'il y a intervention).

Le style d'attachement est considéré comme un facteur de vulnérabilité important associé à diverses formes de psychopathologies comme la dépression, les troubles de l'anxiété, et en particulier avec le trouble de personnalité limite (Mikulincer et Shaver, 2012). Ces mêmes psychopathologies sont généralement associées à un déficit ou à un dysfonctionnement de la capacité de mentalisation, ainsi qu'à des différences significatives dans la structure et l'activité du réseau du mode par défaut (Coutinho *et al.*, 2016 ; Fonagy et Luyten, 2009 ; Luyten et Fonagy, 2018 ; Taubner *et al.*, 2011 ; Yan *et al.*, 2019 ; Yang *et al.*, 2016). Ce n'est pas surprenant, puisque le lien général entre l'attachement et la psychopathologie est conçu comme étant dû à des croyances dysfonctionnelles à propos de soi-même et d'autrui, ainsi qu'à des disruptions du développement des capacités de régulation des émotions, et les problèmes dans les relations interpersonnelles qui en résultent (Mikulincer et Shaver, 2012).

Durant les dernières décennies, les travaux de l'anthropologue Sarah Hrdy (1999, 2000, 2009, 2017) ont permis de montrer que dans le cas de l'espèce humaine, les donneurs

de soins qui s'occupent des enfants ne se limitent pas aux parents, mais comprennent en réalité l'ensemble des membres d'un groupe social élargi composé d'individus qui ne présentent souvent pas de lien de parenté direct. Hrdy capture ce phénomène à travers le concept *d'allop parentalité*, et défend l'hypothèse que le type de collaboration interpersonnelle qu'il implique aurait joué un rôle central dans le développement des habiletés sociocognitives qui sont propres à l'espèce humaine (Hrdy, 2009). En continuité avec cela, il a été récemment proposé que les mécanismes de bas niveau qui sous-tendent le phénomène de l'attachement pourraient servir de bases ontogénétiques pour l'émergence de certaines formes de cognitions sociales typiquement humaines, à travers un processus de coévolution génétique, épigénétique, et culturelle (Granqvist, 2020). Cette hypothèse est en directe continuité avec le modèle développemental proposé dans le cadre de la théorie de la mentalisation, qui affirme que la capacité de mentalisation est acquise à travers les relations d'attachement qui unissent un individu et ses donneurs de soins primaires durant l'ontogénie. À l'aide du travail réalisé jusqu'à présent, nous avons maintenant tous les outils conceptuels en main pour reformuler le modèle développemental de l'acquisition de la capacité de mentalisation proposé par la théorie de la mentalisation dans le cadre du FEP et des approches prédictives du cerveau humain. Les processus et les mécanismes que cette reformulation nous aura permis d'explicitier vont constituer la base de notre explication de la psychothérapie psychodynamique.

3.2.3 Une reformulation du processus développemental d'acquisition de la capacité de mentalisation à travers les relations d'attachement

Le modèle développemental proposé dans cette section servira de base pour notre reconceptualisation générale du processus de changement psychologique qui s'opère à

travers la relation thérapeutique entre un client et son psychothérapeute dans le cadre de la psychothérapie psychodynamique.

Le point de départ du processus de développement de la capacité de mentalisation correspond à la mise en place de la relation d'attachement entre l'enfant et ses donneurs de soins primaires (Fonagy *et al.*, 2018). Nous partons de la prémisse selon laquelle l'enfant et ses donneurs de soins présentent tous les deux une motivation intrinsèque à la réalisation de l'allostase de l'enfant (Cittern *et al.*, 2018), qui dans le cadre des approches prédictives, peut-être comprise comme le processus dynamique de minimisation des erreurs de prédiction intéroceptives de ce dernier (Barrett, 2016 ; Fotopoulou et Tsakiris, 2017). Étant chacun dotés de priors adaptatifs visant à faciliter cette tâche, correspondant respectivement aux systèmes d'attachement et de caregiving (Bowlby, 1999 ; George et Solomon, 2008), l'enfant et ses donneurs de soins présentent une tendance commune à s'engager de manière récurrente, et ce tout au long de l'ontogénie, dans des interactions collaboratives de modulation des états homéostatiques et affectifs de l'enfant (Fonagy *et al.*, 2018). Le modèle proposé par Vasil *et al.* (2020) nous permet de reconceptualiser ces interactions comme des instances de ce que nous considérons comme la première forme de *communication coopérative* à laquelle participe un enfant humain. Selon nous, l'objectif de cette forme primaire de communication coopérative est la minimisation des erreurs de prédictions intéroceptives de l'enfant à travers la modélisation prédictive de ses états affectifs, incluant à la fois des inférences perceptuelles et actives visant à respectivement à les catégoriser et à les réguler.

Au sein de cette pratique communicationnelle, l'enfant et ses donneurs de soins s'engagent dans des *cycles d'action-perception couplés* au sein desquels ils tentent de prédire mutuellement leurs comportements respectifs en implémentant des régimes d'inférence active particuliers (Vasil *et al.*, 2020). En effet, pour chacun des participants, les données sensorielles générées par ces comportements constituent le

meilleur moyen de suivre et de réaliser le processus de minimisation des erreurs de prédiction intéroceptives de l'enfant. À travers des inférences actives de communication affective, comme pleurer ou donner des coups de pied par exemple, l'enfant cherche à prédire (à causer) les interventions de régulation homéostatique et affective de ses donneurs de soins, qui sont les seuls moyens efficaces qu'il a (au départ) de minimiser ses erreurs de prédiction intéroceptive (Fotopoulou et Tsakiris, 2017). À travers des inférences actives de soin *et* de communication affective, comme nourrir l'enfant *et* lui parler par exemple (Fonagy *et al.*, 2018), les donneurs de soins cherchent à prédire les expressions affectives de l'enfant, qui sont sa meilleure source d'information sur l'état homéostatique et affectif de ce dernier. En effet, ces expressions affectives lui communiquent le résultat du processus d'inférence perceptuelle intéroceptive réalisé par le système prédictif de l'enfant (Fotopoulou et Tsakiris, 2017), et signalent par conséquent *l'incertitude ressentie* par ce dernier quant au maintien de ses états internes dans les normes qui lui sont dictées par son phénotype (c.-à-d., ses valeurs homéostatiques préférées ; Solms et Friston, 2018)).

Selon Vasil *et al.* (2020), le type de communication coopérative décrit ci-dessus serait parallèlement motivé et régulé par un prior adaptatif spécifique aux membres de l'espèce humaine : une prédisposition à collecter des preuves selon lesquelles leurs états mentaux et ceux de leurs congénères seraient alignés ou similaires. À travers le mécanisme de *l'attention jointe* (Tomasello et Carpenter, 2007) — et par extension, de modulation de la précision (Feldman et Friston, 2010) — la tendance à collecter des preuves pour ce prior devrait mener à l'installation de *régimes d'attention partagée* (Ramstead *et al.*, 2016) au sein desquels l'enfant et ses donneurs de soins primaires baseraient leur activité de modélisation prédictive des états affectifs de l'enfant sur les mêmes indices sensoriels. La structuration de l'activité de communication coopérative à travers ces régimes d'attention partagée devrait mener à une synchronisation progressive de l'activité neurocognitive de l'enfant et de ses donneurs de soins primaires, et à la convergence vers un modèle prédictif commun des états affectifs de

l'enfant (Friston et Frith, 2015). Dans ce contexte, la beaucoup plus grande précision des priors entretenus par ses donneurs de soins va mener à un entraînement asymétrique des priors de l'enfant vers ces derniers (Vasil *et al.*, 2020). Autrement dit, l'enfant va être progressivement entraîné vers une adoption généralisée des modèles prédictifs à travers lesquels ses donneurs de minimisent ses erreurs de prédiction intéroceptives. Parmi ces modèles se trouvent des priors empiriques qui expliquent et prédisent les états affectifs et les comportements de l'enfant en les conceptualisant comme étant causés par des états de fait globaux et unitaires qui caractérisent l'échelle cognitive de son organisation : des états mentaux (ex. émotion ; Barrett, 2016 ; Fonagy *et al.*, 2018).

Tel que nous l'avons proposé précédemment, cette activité de *conceptualisation métacognitive* correspondrait à l'exercice des différentes modalités de la capacité de mentalisation (ex. implicite et automatique versus explicite et contrôlée ; Fonagy *et al.*, 2018), une suite d'habiletés qui guide selon nous la grande majorité des régimes d'inférence active implémentés par les donneurs de soins durant les interactions de régulation homéostatique et affective de l'enfant. Puisque les priors qu'elle implémente présentent une précision élevée, l'activité de conceptualisation métacognitive pratiquée par les donneurs de soins devrait déterminer la trajectoire des régimes d'attention partagée qui structurent leurs interactions de communication coopérative avec l'enfant. Ce faisant, les cycles d'action-perception couplés qui définissent ces interactions peuvent être considérés comme étant principalement caractérisés par une activité de conceptualisation métacognitive visant à la modélisation erreurs de prédiction intéroceptives de l'enfant. La synchronisation des dynamiques d'organisation neurocognitive de l'enfant à celle de ses donneurs de soins passerait donc par un processus de conditionnement à travers lequel ces derniers l'entraîneraient à l'exercice de leurs habiletés de conceptualisation métacognitive. L'apprentissage graduel de cette suite d'habiletés correspondrait à l'émergence progressive d'une nouvelle échelle d'organisation de la hiérarchie prédictive incarnée par le cerveau de l'enfant, qui pourrait être identifiée au développement progressif du réseau du mode par défaut

(Barrett et Satpute, 2013 ; Carhart-Harris et Friston, 2010 ; Lindquist et Barrett, 2012 ; Luyten et Fonagy, 2015). Cette hypothèse est appuyée par la recherche empirique, qui montre que le DMN n'existe pas à la naissance, et que la connectivité fonctionnelle qui lie les nœuds qui le composent semble résulter d'un processus développemental dépendant de l'expérience (Gao *et al.*, 2011, 2013).

L'activité de ce nouveau niveau d'organisation neurocognitif consiste à générer et implémenter des concepts métacognitifs, des modèles prédictifs d'un niveau d'abstraction élevée qui expliquent les dynamiques du fonctionnement affectif et comportemental (c.-à-d. émotionnel) de l'enfant et de ses congénères en modélisant les processus cognitifs qui en sont responsables (Barrett et Satpute, 2013 ; Carhart-Harris et Friston, 2010 ; Luyten et Fonagy, 2015). Le complexe formé par l'intégration et la collaboration dynamique de ces concepts métacognitifs dans la modélisation du processus de réalisation de l'allostase de l'enfant mène à la génération du concept de Soi psychologique, un modèle prédictif de son propre fonctionnement cognitif, implémentant des inférences perceptuelles et actives qui expliquent et dirigent son activité cognitivo-comportementale afin de favoriser l'accomplissement de ses besoins biologiques à travers des interactions adaptatives avec son environnement social. En effet, le développement de ses habiletés de conceptualisation métacognitives va mener chez l'enfant à l'émergence d'une nouvelle forme d'agentivité sociocognitive qui va l'ouvrir à un nouveau *paysage d'affordances*, des possibilités d'inférence active lui permettant d'optimiser le processus de réalisation de son allostase tout en implémentant un couplage structurel avec une échelle spécifique de sa niche écologique, celle qui correspond aux états mentaux, c'est-à-dire, à l'activité cognitive, de ses congénères (Bruin *et al.*, 2018 ; Vasil *et al.*, 2020). Les possibilités d'inférence active qui s'offrent à l'enfant au sein de ce paysage d'affordances particulier peuvent être organisées en deux catégories principales : a) des inférences actives visant à la

prédiction de sa propre activité cognitive et comportementale afin d'optimiser le processus de réalisation de son allostase, et que l'on peut regrouper sous la catégorie générale des processus d'autorégulation émotionnelle ; et b) des inférences actives visant à la prédiction de l'activité cognitive et comportementale de ses congénères afin d'instaurer et un couplage structurel robuste avec son économe socioculturelle, et que l'on peut classer dans la catégorie générale de la cognition sociale.

C'est au niveau d'organisation des systèmes neurocognitifs humains correspondant à l'accomplissement des fonctions d'autorégulation émotionnelle et de cognition sociale à travers l'exercice des habiletés de conceptualisation métacognitive que nous proposons d'identifier l'espace d'échelle correspondant aux phénomènes en jeu au sein de la pratique de la psychothérapie psychodynamique. En effet, d'après la psychologie psychodynamique, les émotions et les relations interpersonnelles sont les deux facteurs les plus importants pour la santé mentale d'un individu. De ce fait, nous proposons que l'un des objectifs principaux de la psychothérapie psychodynamique puisse être défini comme l'entraînement du client à l'exercice habile (fonctionnel) des inférences actives d'autorégulation émotionnelle et de cognition sociale. De plus, nous faisons l'hypothèse que le processus de changement psychologique qui s'opère par le biais de la relation entre un patient et son psychothérapeute dans le cadre de la pratique de la psychothérapie psychodynamique repose sur les mêmes mécanismes que le processus d'apprentissage des habiletés de conceptualisation métacognitive à travers la relation d'attachement entre un enfant et ses donneurs de soins primaires. Ainsi, sur la base du travail réalisé jusqu'ici, nous formulons l'hypothèse selon laquelle la psychothérapie psychodynamique peut être définie comme une pratique culturelle partagée au sein de laquelle un donneur de soin aux caractéristiques particulières (un professionnel entraîné) assiste et supervise un individu dans l'apprentissage d'habiletés de conceptualisation métacognitive optimales, incluant les concepts

neurocognitifs particuliers qu'elles implémentent, et ce dans le but de favoriser l'amélioration des capacités d'autorégulation émotionnelle et de cognition sociale qui seraient essentielles à sa santé mentale. Avant pouvoir procéder à une reconceptualisation détaillée de cette pratique, la dernière étape va consister à expliciter certaines conditions qui sous-tendent sa possibilité et son efficacité.

3.2.4 Les soins cognitifs et le prior de confiance épistémique

À travers l'ontogénie, l'enfant va développer une autonomie de plus en plus grande en ce qui concerne l'implémentation des inférences actives proximales nécessaires à la réalisation de son allostasie. Si l'on prend le domaine de l'alimentation comme exemple, bien que ses donneurs de soins doivent encore lui fournir sa nourriture jusqu'à sa maturité complète, l'enfant va progressivement devenir capable de s'alimenter lui-même, puis de cuisiner lui-même sa nourriture, et ainsi de suite. De même, une fois ses habiletés de conceptualisation métacognitive suffisamment développées, l'enfant devrait devenir capable de réguler son activité cognitive afin de générer des comportements lui permettant de s'adapter de manière minimalement efficace à sa niche écologique. Cela implique qu'au fur et à mesure du développement, le rôle des donneurs de soins dans la réalisation de l'allostasie de l'enfant va se modifier et se restreindre progressivement pour prendre la forme d'interventions de plus en plus indirectes. Plus précisément, à partir du développement d'un certain niveau de maturité chez l'enfant, les soins qui lui sont procurés par ses donneurs de soins vont se limiter en grande partie à des interventions de régulation de son activité cognitive et comportementale par le biais de la communication, langagière principalement. Ces interventions correspondent à des inférences actives visant à influencer l'activité cognitive de l'enfant de manière à ce qu'elle se traduise en des comportements

adaptatifs qui favorisent la réalisation de son allostasie. Si nous reprenons l'exemple de l'alimentation, les donneurs de soins ne vont plus directement alimenter l'enfant, mais vont lui verbaliser la directive selon laquelle il doit manger. Ces interventions constituent une extension de l'activité de co-construction conceptuelle correspondant à la mentalisation explicite. Dans le cas de l'alimentation, il s'agit de solliciter l'enfant à exercer un contrôle exécutif sur son activité cognitive de manière à ce qu'il implémente des inférences actives permettant la réalisation efficace de son allostasie.

Autrement dit, les inférences actives des donneurs de soins vont progressivement se spécialiser vers la prédiction (la régulation) de l'activité cognitive de l'enfant, servant de moyen intermédiaire pour participer à la réalisation de son allostasie. Dans ce contexte, les interactions entre un enfant et ses donneurs de soins vont prendre la forme d'une pratique culturelle partagée de communication coopérative telle que décrite originellement par Vasil *et al.* (2020). Plus précisément, cette pratique va être principalement caractérisée par une activité de co-conceptualisation métacognitive de l'activité cognitive de l'enfant, et plus particulièrement de la co-construction des concepts émotionnels à travers lesquels il réalise son allostasie. Cette activité devrait normalement couvrir l'intégralité des concepts complexes à l'aide desquels l'enfant explique et régule son expérience affective et ses comportements, allant jusqu'à des concepts relativement abstraits comme ceux qui portent sur sa personnalité ou son identité.

Relativement à la tâche de réalisation complète de l'allostasie de l'enfant qui leur incombait auparavant, les interventions de régulation cognitive décrites ci-dessus présentent des couts énergétiques et métaboliques très faibles pour ses donneurs de soins primaires. Selon nous, cette baisse du cout des soins nécessités par l'enfant est l'une des causes principales de l'élargissement drastique de l'ensemble des donneurs de soins potentiels qui s'opère à partir d'un certain niveau de son développement. En effet, et tel que le concept d'aloparentalité présenté précédemment le souligne, les

enfants humains jouissent d'une quantité élevée de donneurs de soins potentiels (Hrdy, 2009). Nous pensons que cette abondance résulte en partie du fait que la majorité de ces donneurs de soins ne présentent pas un niveau de responsabilité et d'investissement élevé vis-à-vis de la tâche de réalisation de l'allostasie de l'enfant. Plus précisément, nous proposons que si ces *donneurs de soins secondaires* sont motivés à s'occuper d'un enfant, ce soit par ce que leur rôle est principalement limité à la tâche relativement peu coûteuse de régulation du fonctionnement cognitif de ce dernier. De manière heuristique, nous pensons qu'il ne serait pas injustifié de qualifier ces derniers de *donneurs de soins cognitifs*. Cependant, étant donné la spécialisation de l'espèce humaine dans des formes de cognition complexes, et l'importance de ces dernières dans son mode de vie caractéristique, ces donneurs de soins jouent selon nous un rôle extrêmement important dans le développement d'un individu.

Dans le cadre des niches socioculturelles complexes qui caractérisent les sociétés modernes dans lesquelles nous vivons, le rôle central de ce type de soins cognitifs a donné lieu à l'émergence de pratiques culturelles institutionnalisées, qui exercent une fonction régulatrice essentielle sur les populations humaines. Parmi ces pratiques, la première et la plus importante correspond bien évidemment à l'éducation, dont les acteurs principaux constituent les enseignants et les professeurs. Ces donneurs de soins particuliers sont spécialisés dans le développement des habiletés cognitives reliées à l'acquisition de connaissances, à la fois abstraites et pratiques, favorisant l'adaptation et l'intégration d'un individu aux niches socioculturelles complexes typiques à l'espèce humaine. Une autre pratique culturelle institutionnalisée, développée plus récemment quant à elle, correspond à celle qui fait l'objet de ce mémoire de maîtrise : la psychothérapie. Selon nous, et bien que se limitant à la régulation du fonctionnement cognitif d'un individu, les soins offerts par un psychothérapeute devraient être du même type que ceux qui lui sont fournis par ses donneurs de soins primaires. Plus précisément, nous pensons que les psychothérapeutes correspondent à un type particulier de donneurs de soins dont le rôle consiste à assister l'individu dans la tâche

de conceptualisation métacognitive des processus cognitifs qui régissent son fonctionnement psychologique ou mental. Avant de pouvoir procéder à la synthèse de notre reconceptualisation de la psychothérapie psychodynamique, il nous reste à clarifier l'un des mécanismes centraux qui sous-tendent la possibilité du couplage structurel neurocognitif intime qui peut s'opérer entre un patient et son psychothérapeute, et qui est nécessaire pour que cette pratique puisse initier un changement significatif de son fonctionnement cognitif.

Comme nous l'avons vu précédemment, l'extrême dépendance des enfants humains à leurs donneurs de soins primaires pour la réalisation de leur allostase les mène à développer des priors empiriques, les schémas d'attachement, dont la fonction est d'évaluer le niveau de confiance relative qu'ils peuvent accorder au support potentiel des autres individus dans la tâche de régulation de leurs états homéostatiques et affectifs. Dans cette section, nous avons vu qu'à partir d'un niveau de maturité suffisant, la dépendance de l'enfant à ses donneurs de soins se retreint en grande partie à la régulation du processus de conceptualisation métacognitive à travers lequel il implémente lui-même la réalisation de son allostase. Étant donné le rôle crucial que joue ce type de soins dans le développement cognitif de l'enfant, ne serait-il pas adaptatif pour lui de développer un modèle interne qui, de manière similaire aux schémas d'attachements, servirait à optimiser son rapport à ses donneurs de soins cognitifs ? Dans cet ordre d'idée, il a été proposé que le manque de confiance accordé à l'environnement social proximal qui est associé aux styles d'attachement insécures et ambivalents puisse mener à une forme *d'hypervigilance épistémique* dans laquelle l'individu adopte un scepticisme général vis-à-vis de l'information qui lui est véhiculée lors des interactions interpersonnelles (Fonagy *et al.*, 2015 ; Sperber *et al.*, 2010). En continuité avec cela, Fonagy et Allison (2014), ont avancé le concept de *confiance épistémique* pour capturer l'ouverture relative d'un individu aux informations qui lui sont communiquées par autrui dans le cadre des relations interpersonnelles.

En suivant la formule avancée par Cittern pour redéfinir les schémas d'attachement, nous proposons que dans le cadre des approches prédictives, la notion de confiance épistémique puisse être reconceptualisée comme un prior empirique qui encode le niveau de confiance accordée par un individu au support potentiel de ses congénères dans la tâche de régulation de son fonctionnement cognitif. Telle que nous l'avons décrite précédemment, cette tâche de régulation passe normalement par une pratique de communication coopérative au sein de laquelle ses congénères influencent son activité de conceptualisation métacognitive à travers l'installation de régimes d'attention partagée. Dans ce contexte, le prior de confiance épistémique devrait se manifester par une modulation du niveau de précision accordée aux informations qui lui sont véhiculées à travers les inférences actives de communication qui sont déployées par ses congénères, ce qui détermine l'influence qu'elles auront sur la modification des concepts neurocognitifs implémentés par le système prédictif de l'individu.

Étant donné le lien intime qui lie l'attention et la précision dans le cadre des approches prédictives, les régimes d'attention partagée implémentés par les donneurs de soins peuvent être considérés comme des régimes de précision partagée. Puisque la précision est accordée à travers l'investissement des ressources attentionnelles, alors lorsqu'un individu participe à un régime d'attention partagée, il s'ouvre à la possibilité d'une modification de ses priors à travers la modulation de la précision qui leur est accordée. Dans ce contexte, nous proposons que, pour être efficace, le prior de confiance épistémique devrait non seulement déterminer le niveau de précision accordé à l'information qui lui est véhiculée au sein des régimes d'attention partagée auxquels il participe, mais aussi à réguler la propension d'un individu à s'engager dans ces derniers. Puisque les régimes d'attention partagée sont implémentés à travers le mécanisme de l'attention jointe, nous faisons l'hypothèse selon laquelle le prior de confiance épistémique devrait être implémenté en partie par un processus d'inhibition sélective de l'attention jointe permettant à un individu de choisir (implicitement ou explicitement) les régimes d'attention partagée dans lesquels il souhaite s'engager.

Nous pensons que l'introduction du prior de confiance épistémique peut nous permettre d'appuyer l'idée selon laquelle l'activité de conceptualisation métacognitive correspond bien aux processus qui définissent une échelle d'auto-organisation particulière des systèmes neurocognitifs humains. Plus précisément, nous pensons que le prior de confiance épistémique peut-être conçu comme ayant pour fonction d'implémenter la clôture opérationnelle d'un organisme humain dans le cadre de son couplage structurel avec son éconiche socioculturelle. Dans ce contexte, nous proposons que la génération du prior de confiance épistémique puisse être conçu comme le mécanisme clé permettant l'émergence d'une couverture de Markov qui aurait pour fonction d'opérer une différenciation entre les états internes et les états externes de l'organisme à une échelle spécifique de son organisation : celle qui détermine son fonctionnement en tant qu'agent cognitif. Ce faisant, le prior de confiance épistémique encode selon nous une croyance métacognitive qui régule les échanges cognitifs qui s'opèrent entre un agent cognitif humain et ses congénères.

De manière très intéressante pour nous, Fonagy et Allison (2014) ont utilisé le concept de confiance épistémique pour proposer un modèle explicatif de l'installation de l'alliance thérapeutique, qui est l'un des facteurs considérés comme le plus important pour prédire l'efficacité thérapeutique de n'importe quelle forme de psychothérapie (Baldwin *et al.*, 2007 ; Crits-Christoph *et al.*, 2011 ; Falkenström et Larsson, 2017 ; Flückiger *et al.*, 2011, 2020). L'alliance thérapeutique est définie comme le degré d'entente qui détermine la capacité d'un client et de son psychothérapeute à collaborer efficacement vers la réalisation des buts qu'ils se sont fixés ensemble (Ardito et Rabellino, 2011). Dans le cadre de la théorie de la mentalisation, Fonagy et Allison (2014) ont proposé que l'installation d'une bonne alliance thérapeutique dépende de la capacité de mentalisation du thérapeute, qui lui permettrait de développer et de communiquer au client une compréhension adéquate de son expérience affective et de son fonctionnement cognitif, ce qui devrait mener ce dernier à développer un certain niveau de confiance dans la capacité du thérapeute à comprendre ses états mentaux.

Cette confiance graduellement accordée au thérapeute va ouvrir peu à peu le client à l'information qu'il lui véhicule, et le mènerait à accepter son influence et à collaborer avec lui dans la tâche d'interprétation et de régulation de ses états mentaux. Une fois le client ouvert à l'influence du psychothérapeute sur son fonctionnement psychologique, Fonagy et Allison (2014) affirment que l'étape suivante consiste simplement en l'engagement de ceux-ci dans une activité collaborative de mentalisation de l'activité cognitive et comportementale du patient. La pratique récurrente de cette activité aurait pour conséquence l'amélioration progressive des capacités de mentalisation du patient, ainsi qu'une modification substantielle de son niveau de confiance épistémique général vis-à-vis de ses congénères. La dernière étape, qui serait critique pour l'efficacité thérapeutique de la psychothérapie, correspondrait à la généralisation de cette attitude d'ouverture épistémique à l'ensemble des relations interpersonnelles entretenues par l'individu à l'extérieur de la psychothérapie. Cela aurait pour conséquence de permettre à l'individu d'apprendre des expériences vécues dans le cadre de ses relations interpersonnelles, de manière à améliorer d'autant plus ses capacités de mentalisation, ainsi que des habiletés d'autorégulation émotionnelle et de cognition sociales qui en résultent.

La position exposée ci-dessus nous permet d'explicitier un aspect problématique de la reformulation du concept de confiance épistémique proposée précédemment. En effet, l'inhibition des informations provenant d'autrui et du mécanisme de l'attention jointe semble à première vue s'opposer toutes deux à la collecte de preuves pour le prior d'alignement des états mentaux tel que définit par Vasil et al. 2020. Cependant, Fonagy et Allison nous fournissent directement une piste de réponse à cette problématique, en proposant que ce soit la capacité de mentalisation du psychothérapeute qui, en montrant à l'individu que ses états mentaux sont compris par autrui, permettrait la modification du niveau de confiance épistémique qui lui est accordé. Cette proposition nous permet de réaliser que, dans le cas d'une incompatibilité relative des priors entretenus par les participants, l'activité de communication coopérative peut aussi mener à la génération

de preuves qui iraient à l'encontre du prior d'alignement. Ainsi, pour un individu qui aurait du mal à minimiser les erreurs de prédictions reliées à son fonctionnement mental et émotionnel, la communication coopérative peut être considérée comme une activité risquée, c'est-à-dire, pleine d'incertitude. Dans ce contexte, on peut comprendre que le prior de confiance épistémique ne contredit pas nécessairement le prior d'alignement, mais qu'au contraire, il peut permettre l'identification et l'engagement sélectif dans les instances de communication coopérative qui vont permettre de générer des preuves pour ce dernier.

Ainsi, il est possible de reconceptualiser le rôle des habiletés de conceptualisation métacognitive du thérapeute dans la modification du prior de confiance épistémique comme étant la démonstration de sa capacité à générer des preuves pour le prior d'alignement. Cette démonstration aurait pour conséquence de réduire l'incertitude ressentie par le client vis-à-vis de son engagement potentiel dans une activité de communication coopérative avec son psychothérapeute. Cela devrait mener le client à accepter de participer aux régimes d'attentions partagée proposés par le psychothérapeute, ce qui revient à laisser à ce dernier l'opportunité d'influencer son activité de conceptualisation métacognitive en modulant la précision accordée aux prédictions et aux erreurs de prédictions reliées aux concepts neurocognitifs qui régissent son fonctionnement émotionnel et interpersonnel. En résumé, nous proposons que la modification du prior de confiance épistémique correspond à un mécanisme de déclenchement qui détermine la possibilité d'établir le type de synchronisation neurocognitive (Friston et Frith, 2015) nécessaire à la modification du fonctionnement mental d'un individu qui est recherché dans le cadre de la psychothérapie. Plus spécifiquement, nous considérons le prior de confiance épistémique comme une forme de paramètre du modèle interne généré par le cerveau d'un individu, et qui détermine le degré d'intimité du couplage structurel neurocognitif qui peut s'installer entre lui et ses congénères. L'explicitation du rôle déclencheur de la modification du prior de confiance épistémique dans l'instauration du couplage structurel neurocognitif qui peut

s'installer entre un individu et ses donneurs de soins cognitifs nous offre la dernière pièce manquante afin de nous constituer un modèle explicatif suffisamment exhaustif de la pratique de la psychothérapie psychodynamique. Nous pouvons maintenant passer à l'élaboration de ce modèle en procédant à la synthèse du travail réalisé dans ce chapitre.

3.3 La psychothérapie psychodynamique reconceptualisée

Dans cette section finale de notre mémoire de maîtrise, nous allons procéder à une synthèse des modèles théoriques présentés jusqu'à présent, ainsi que des hypothèses qu'ils nous ont permis de formuler, afin de proposer une reconceptualisation des mécanismes qui sous-tendent le fonctionnement et l'efficacité de la pratique de la psychothérapie psychodynamique. Cette synthèse sera structurée en deux étapes, correspondant respectivement à une tentative de réponse aux deux problématiques formulées dans notre introduction. Dans chacune de ces étapes, nous procéderons du général au particulier, en proposant tout d'abord une reconceptualisation globale de l'espace d'échelle de l'organisation des systèmes neurocognitifs humains adressée par la psychothérapie psychodynamique, ainsi que de la manière dont elle tente de l'influencer, puis en formulant des hypothèses plus fines sur les processus et les mécanismes qui pourraient sous-tendre ce niveau d'organisation et sa modification. Ainsi, le modèle proposé ici sera principalement composé d'un ensemble d'hypothèses formulées afin de guider la recherche empirique sur le fonctionnement et l'efficacité thérapeutique de la psychothérapie psychodynamique. Dans l'éventualité où ces hypothèses sont effectivement investiguées empiriquement, et si certaines d'entre elles s'avèrent validées, elles pourront servir de base pour l'élaboration d'un nouveau cadre

théorique permettant de raffiner ou même de réformer la pratique de la psychothérapie psychodynamique afin de renforcer son efficacité thérapeutique.

3.3.1 Le niveau organisation des systèmes neurocognitifs humains adressé par la psychothérapie psychodynamique

- À quel(s) niveau(x) d'organisation spécifique(s) des systèmes neurocognitifs humains correspondent les phénomènes impliqués dans la pratique de la psychothérapie psychodynamique ? Quels sont les mécanismes ou les processus qui définissent l'organisation de ce(s) niveau(x) ?

La pratique de la psychothérapie psychodynamique repose sur l'engagement avec un niveau particulier de l'organisation des systèmes neurocognitifs humains (Carhart-Harris et Friston, 2010 ; Connolly et van Deventer, 2017 ; Ramstead *et al.*, 2017), dont l'existence et le fonctionnement sont déterminés par le couplage structurel intime qui lie un être humain avec son éconiche socioculturelle (Bruin *et al.*, 2018 ; Ramstead *et al.*, 2016), et qui s'instaure par le biais du processus collaboratif de réalisation de l'allostase d'un individu à travers les relations d'attachement qui l'unissent avec ses donneurs de soins primaires durant l'ontogénie (Atzil *et al.*, 2018 ; Atzil et Barrett, 2017 ; Cittern *et al.*, 2018 ; Fotopoulou et Tsakiris, 2017). Plus précisément, les interactions récurrentes de communication coopérative visant à la modulation des états homéostatiques et affectifs d'un enfant humain mènent, à travers la synchronisation progressive des systèmes prédictifs incarnés par les cerveaux de ce dernier et des adultes qui s'en occupent, à l'émergence chez celui-ci d'une nouvelle échelle d'organisation neurocognitive qui implémente la génération d'un modèle commun du phénotype et de la niche écologique qu'il partage avec ses congénères (Friston et Frith,

2015 ; Vasil *et al.*, 2020). *Les processus cognitifs qui définissent ce niveau d'organisation correspondent à une suite d'habiletés de conceptualisation métacognitive (de mentalisation) offrant à l'individu la capacité de modéliser les processus causaux qui définissent l'échelle cognitive de l'organisation et de l'activité des organismes humains* (Barrett, 2016 ; Fonagy *et al.*, 2018 ; Moutoussis *et al.*, 2014).

L'apprentissage de ces habiletés est sous-tendu par le développement des échelles supérieures de la hiérarchie prédictive incarnée par le cerveau humain, et en particulier du *réseau du mode par défaut*, dont l'activité génère et implémente des concepts métacognitifs, des modèles prédictifs d'un niveau d'abstraction élevée qui expliquent les dynamiques du fonctionnement affectif et comportemental (c.-à-d. émotionnel) de l'individu et de ses congénères en modélisant les processus cognitifs qui en sont responsables (Barrett et Satpute, 2013 ; Carhart-Harris et Friston, 2010 ; Luyten et Fonagy, 2015). Le complexe formé par l'intégration et la collaboration dynamique de ces concepts métacognitifs dans la modélisation du processus de réalisation de l'allostasie de l'individu mène à la génération du *concept de Soi psychologique*, un modèle prédictif de son propre fonctionnement cognitif, implémentant des inférences perceptuelles et actives qui expliquent et dirigent son activité cognitivo-comportementale afin de favoriser l'accomplissement de ses besoins biologiques à travers des interactions adaptatives avec son environnement social.

En effet, l'exercice de ses habiletés de conceptualisation métacognitives sous-tend une forme d'agentivité sociocognitive spécifique au sein de laquelle un être humain peut s'engager dans un paysage d'affordances particulier, composé de possibilités d'inférences actives lui permettant de réaliser son allostasie en implémentant un couplage structurel intime avec une échelle spécifique de sa niche écologique, celle qui correspond aux états mentaux, c'est-à-dire à l'activité cognitive, de ses congénères (Bruin *et al.*, 2018 ; Bruineberg et Rietveld, 2014 ; Vasil *et al.*, 2020). Le niveau d'organisation caractérisé par l'exercice des habiletés de conceptualisation

métacognitive implémente deux fonctions principales, caractérisées par deux catégories d'inférences actives spécifiques, qui correspondent aux deux mécanismes principaux à travers lesquels il contribue à la minimisation de l'énergie libre des organismes humains :

- 1- L'autorégulation émotionnelle : l'implémentation de concepts émotionnels (Barrett, 2016), des modèles prédictifs multisensoriels qui minimisent les erreurs de prédictions intéroceptives, extéroceptives et proprioceptives du système prédictif en générant des prédictions qui les expliquent comme résultant de causes distales spécifiques, les propriétés (processus, états, etc.) qui caractérisent les régularités de l'organisation neurocognitive du système lui-même dans son interaction avec sa niche écologique à différentes échelles spatiotemporelles (émotions, personnalité, soi, etc.). Incarnées par des signaux neuronaux modulateurs descendants, ces prédictions implémentent un contrôle exécutif qui régule et dirige l'activité des échelles intermédiaires de la hiérarchie cérébrale qui sont responsables de l'expérience affective ainsi que de l'activité cognitive et comportementale du système. Les inférences actives d'autorégulation émotionnelle constituent l'un des moyens principaux à travers lesquels les organismes humains optimisent le processus de réalisation de leur allostase (Barrett, 2016). Autrement dit, elles correspondent à l'une des manières particulières que les systèmes neurocognitifs humains ont trouvées pour optimiser le processus dynamique de maintien de leurs états internes dans les normes homéostatiques définies par leur phénotype, ce qui revient à minimiser leur énergie libre variationnelle (Ramstead *et al.*, 2017).
- 2- La cognition sociale : l'implémentation de concepts mentaux (émotions, intentions, etc.), des modèles prédictifs multisensoriels qui minimisent les erreurs de prédictions liées aux comportements d'autrui lors des interactions sociales en générant des prédictions qui les expliquent comme résultant de causes distales spécifiques, les processus cognitifs qui en responsables

(Moutoussis *et al.*, 2014b). Les inférences actives de cognition sociale constituent l'un des moyens principaux à travers lesquels les organismes humains optimisent le processus d'adaptation à leur environnement social, dont les membres d'une espèce sociale comme l'espèce humaine dépendent pour la réalisation de leur allostasie (Atzil *et al.*, 2018 ; Atzil et Barrett, 2017). Autrement dit, elles correspondent à l'une des manières particulières que les systèmes neurocognitifs humains ont trouvé pour instaurer et maintenir un couplage structurel intime avec leur éconiche socioculturelle, ce qui revient à s'assurer de maintenir les échanges de ressources nécessaires à la réalisation de leur allostasie, et donc à minimiser leur énergie libre variationnelle (Ramstead *et al.*, 2017).

Compte tenu du rôle crucial que les capacités d'autorégulation émotionnelle et de cognition sociale jouent dans la minimisation de l'énergie libre des systèmes neurocognitifs humains, il n'est pas étonnant que la dysfonction des habiletés de conceptualisation métacognitive qui les sous-tendent puisse mener à des conséquences pathologiques sur le fonctionnement psychologique d'un individu. Ce sont précisément ces formes de psychopathologie que la psychothérapie psychodynamique tente de soigner. *La psychothérapie psychodynamique est une méthode d'intervention psychosociale spécialisée dans le traitement des formes de troubles mentaux qui résultent et/ou sont influencées par le dysfonctionnement des habiletés de conceptualisation métacognitive qui sous-tendent les capacités d'autorégulation émotionnelles et de cognition sociale essentielles à la santé mentale d'un agent cognitif humain.*

Plus spécifiquement, la psychothérapie psychodynamique accorde une attention toute particulière aux types de fonctionnements émotionnels et interpersonnels non adaptatifs qui résultent d'une dominance exagérée de la partie automatique et implicite des processus de conceptualisation métacognitive (Fonagy et Luyten, 2012 ; Solms,

2018). Performés par l'activité relativement isolée des nœuds du réseau du mode par défaut (Luyten et Fonagy, 2015 ; Van Overwalle et Vandekerckhove, 2013), ces processus cognitifs correspondraient à un mode de « pilote automatique » de la fonction de conceptualisation métacognitive (Fonagy et Luyten, 2012 ; Vatansever *et al.*, 2017), qui se manifesterait sur le plan phénoménologique sur le mode de la conscience préreflexive. La prédominance excessive de ce mode de métacognition entrainerait une dynamique d'organisation neurocognitive dans laquelle le système prédictif incarné par le cerveau d'un individu favorise l'évitement proximal de l'augmentation temporaire de l'incertitude ressentie (intensité affective ; Solms et Friston, 2018) au détriment de l'installation d'une forme de métastabilité qui permettrait une meilleure minimisation de l'énergie libre globale du système à des échelles temporelles plus élevées (Friston, 2000 ; Tognoli et Kelso, 2014). Cette dynamique serait implémentée par des régimes d'attention (Ramstead *et al.*, 2016), c'est-à-dire, d'allocation de précision (Clark, 2013), qui favorisent l'implémentation de concepts neurocognitifs dont les prédictions permettent une minimisation immédiate de l'incertitude ressentie (affect) à un niveau tolérable pour le système prédictif. L'augmentation globale de l'énergie libre du système prédictif causée par cette dynamique d'organisation, se manifestant par un fonctionnement émotionnel et interpersonnel non adaptatif, correspond à la forme de détresse psychologique que la psychothérapie psychodynamique tente d'apaiser : *la souffrance émotionnelle* (Solms, 2018).

Afin de traiter la souffrance émotionnelle des individus, la psychothérapie psychodynamique cherche à renforcer la partie explicite et contrôlée des habiletés de conceptualisation métacognitive d'un individu, qui permettent à ce dernier d'exercer un contrôle exécutif sur les processus de conceptualisation métacognitive automatiques afin de les réguler et de modifier les concepts émotionnels et mentaux qu'ils génèrent et qu'ils implémentent (Barrett, 2016 ; Fonagy et Luyten, 2012 ; Solms, 2013, 2018). Performés par l'activité collaborative des nœuds du réseau du mode par défaut et du réseau de contrôle exécutif (Cocchi *et al.*, 2013 ; Luyten et Fonagy, 2015 ; Valk *et al.*,

2016 ; Van Overwalle et Vandekerckhove, 2013), ces processus cognitifs correspondraient à un mode de cognition représentationnelle et langagière impliquant une prise de conscience réflexive de l'activité de conceptualisation métacognitive (Fonagy et Luyten, 2012 ; Solms, 2018). Ce mode de métacognition mènerait à l'engagement sélectif dans une dynamique d'organisation neurocognitive favorisant l'installation d'une forme de métastabilité, qui bien que demandant de tolérer des augmentations ponctuelles de l'incertitude ressentie (intensité affective), permettrait une meilleure minimisation de l'énergie libre globale du système à travers le temps. Ce mode de métacognition accomplirait cela en dirigeant les ressources attentionnelles afin d'allouer plus de précision aux concepts neurocognitifs dont les prédictions implémentent un processus exploratoire visant à l'amélioration des modèles prédictifs qui gouvernent le fonctionnement émotionnel et interpersonnel de l'individu.

3.3.2 La modification de l'organisation des systèmes neurocognitifs humains permise par la pratique de la psychothérapie psychodynamique

- Quels sont les mécanismes à travers lesquels la psychothérapie psychodynamique amène une amélioration du fonctionnement cognitif et de la santé mentale d'un individu ? Comment ces mécanismes affectent-ils l'organisation du ou des niveau(x) des systèmes neurocognitifs humains auxquels ils correspondent ?

La psychothérapie psychodynamique correspond à une tentative de recréer les conditions naturelles de l'apprentissage des habiletés de conceptualisation métacognitives qui sous-tendent les capacités d'autorégulation émotionnelle et de cognition sociale d'un être humain, et ce afin de faciliter leur modification. Ces conditions correspondent à des relations interpersonnelles au sein desquelles des

donneurs de soins cognitifs régulent et façonnent les capacités de conceptualisation métacognitive d'un individu, à travers des interactions récurrentes de communication coopérative qui mènent les dynamiques de l'organisation des échelles supérieures du système neurocognitif de ce dernier à se synchroniser progressivement avec les leurs (Fonagy *et al.*, 2018 ; Vasil *et al.*, 2020). Afin de favoriser l'acquisition de capacités d'autorégulation émotionnelle et de cognition sociale fonctionnelles et adaptatives, la psychothérapie psychodynamique est pratiquée par un donneur de soins cognitifs aux caractéristiques spécifiques, un professionnel ayant suivi une formation²² afin de développer de très bonnes habiletés de conceptualisation métacognitive, et que l'on appelle un psychothérapeute. *La psychothérapie psychodynamique peut être définie comme une pratique culturelle partagée de communication coopérative au sein de laquelle un psychothérapeute et son client s'engagent dans une activité de co-conceptualisation métacognitive des processus cognitifs qui sous-tendent le fonctionnement émotionnel et interpersonnel de ce dernier* (Gendron et Barrett, 2018 ; Vasil *et al.*, 2020). *Cette pratique vise à l'amélioration des habiletés de conceptualisation métacognitive du client, ainsi qu'à entraîner ce dernier à l'exercice habile des capacités d'autorégulation émotionnelle et de cognition sociale qu'elles sous-tendent, tout ceci dans le but de favoriser la réalisation de son allostasie à travers l'instauration d'un couplage structurel robuste et intime avec son éconiche socioculturelle.* Cette modification du fonctionnement cognitif du client passe par plusieurs mécanismes clés.

²² Pour ce qui est des psychothérapeutes d'orientation psychodynamique, cette formation comprend deux volets : a) un apprentissage théorique et pratique leur permettant d'avoir une bonne compréhension du fonctionnement psychologique humain et de savoir l'utiliser dans le contexte des interactions de communication coopérative avec un client ; et b) le suivi d'une psychothérapie personnelle leur permettant de comprendre et de maîtriser leur propre fonctionnement émotionnel et interpersonnel afin qu'il n'exerce pas d'influences néfastes dans le cadre la relation thérapeutique avec un client.

Étant donné le rôle déterminant des soins cognitifs qu'ils reçoivent de leurs congénères dans le développement de leur fonctionnement cognitif, les êtres humains se seraient dotés d'un prior empirique, le prior de confiance épistémique, leur permettant d'évaluer la crédibilité relative qu'il peuvent accorder à des donneurs de soins cognitifs potentiels, et de décider de l'influence relative que ces derniers pourront exercer sur l'organisation du système neurocognitif qu'ils incarnent (Cittern *et al.*, 2018 ; Fonagy et Allison, 2014). L'exercice habile de ses capacités de conceptualisation métacognitives permet au psychothérapeute de développer et de communiquer au client une compréhension adéquate de son expérience affective et de son fonctionnement cognitif et comportemental, ce qui a pour effet de démontrer sa crédibilité en tant que donneur de soins cognitifs potentiel. Cela devrait mener à une modification du prior de confiance épistémique du client dans le cadre de la relation thérapeutique, menant ce dernier à s'ouvrir progressivement à l'influence du thérapeute sur son activité cognitive. Plus précisément, cela va permettre au client de se laisser progressivement entraîner dans la pratique de communication coopérative proposée par le psychothérapeute, la même pratique à travers laquelle le client avait acquis ses habiletés de conceptualisation métacognitive durant son ontogénie.

Au sein de cette pratique de communication coopérative, le client et son psychothérapeute s'engagent dans des cycles d'action-perception couplés au sein desquels ils tentent de prédire et d'expliquer l'activité cognitive et comportementale du client en implémentant des régimes d'inférence active particuliers, composés essentiellement d'inférences actives de communication langagière (Gendron et Barrett, 2018 ; Solms, 2018 ; Vasil *et al.*, 2020). Selon Vasil et al. (2020), ce type de communication coopérative serait motivé et régulé par un prior adaptatif spécifique aux membres de l'espèce humaine : une prédisposition à collecter des preuves selon lesquelles leurs états mentaux et ceux de leurs congénères seraient alignés ou similaires. À travers le mécanisme de l'attention jointe (Tomasello et Carpenter, 2007) — et par extension, de modulation de la précision (Feldman et Friston, 2010) — la

tendance à collecter des preuves pour ce prior devrait mener à l'installation de régimes d'attention partagée (Ramstead *et al.*, 2016) au sein desquels le psychothérapeute et le client baseraient leur activité de modélisation prédictive des processus cognitifs responsables du fonctionnement émotionnel et interpersonnel de ce dernier sur les mêmes indices sensoriels et abstraits²³. La structuration de l'activité de communication coopérative à travers ces régimes d'attention partagée devrait mener à une synchronisation progressive de l'activité neurocognitive du psychothérapeute, et à la convergence vers un modèle prédictif commun du fonctionnement cognitif de ce dernier (Friston et Frith, 2015). Dans ce contexte, la plus grande précision des priors entretenus par le psychothérapeute va mener à un entraînement asymétrique des priors du client vers ces derniers (Vasil *et al.*, 2020). Autrement dit, le client va être progressivement entraîné vers une adoption généralisée des modèles prédictifs à travers lesquels son psychothérapeute explique les erreurs de prédiction reliées à son fonctionnement émotionnel et relationnel.

La manière spécifique à travers laquelle le psychothérapeute est capable de diriger les cycles d'action-perception qui caractérisent la pratique de communication coopérative avec son client, et donc de guider l'activité de conceptualisation métacognitive de ce dernier, consiste à mettre en place des régimes d'attention partagée particuliers, dont une partie importante repose sur l'utilisation des techniques thérapeutiques de la psychothérapie psychodynamique : l'interprétation de l'association libre et du transfert. En effet, ces deux techniques correspondent à l'engagement du psychothérapeute et du client dans une pratique d'attention jointe portant sur l'activité cognitive (association libre) et comportementale (transfert) spontanée du client.

²³ Nous parlons ici de représentations ou de concepts neurocognitifs relativement abstraits issus d'un processus de traitement des informations sensorielles.

La première technique thérapeutique de la psychothérapie psychodynamique, l'association libre, consiste à demander au client de parler de tout ce qui lui vient à l'esprit, en se retenant dans la mesure du possible d'appliquer toute censure morale ou logique à son propos. D'après la compréhension de l'organisation des systèmes neurocognitifs humains qui nous est offerte par le FEP et les approches prédictives, l'activité cognitive spontanée du client devrait être dirigée par l'impératif distal de minimisation de l'énergie libre, et donc par l'impératif proximal de minimisation de l'incertitude ressentie (affect), ce qui devrait se manifester par une activité de modélisation prédictive visant à minimiser les erreurs de prédiction reliées au processus de réalisation de son allostasie (Barrett, 2016 ; Friston et Klaas E., 2007 ; Solms et Friston, 2018). Tel que nous l'avons proposé précédemment, les deux voies principales par le biais desquelles les systèmes neurocognitifs humains optimisent le processus de réalisation de leur allostasie correspondent à des activités d'autorégulation émotionnelle et de cognition sociale à travers l'exercice de leurs habiletés de conceptualisation métacognitive. Cela implique que l'activité cognitive spontanée du client devrait consister majoritairement en un processus de conceptualisation métacognitive portant sur les domaines de son existence qui comportent le plus haut niveau d'incertitude vis-à-vis du processus de réalisation de son allostasie, c'est-à-dire, qui donnent lieu à la plus grande souffrance émotionnelle.

Comme nous le savons, la niche écologique d'un être humain est composée principalement d'autres êtres humains, et le processus de réalisation de son allostasie dépend en grande partie de sa capacité à maintenir un couplage structurel solide avec ses congénères. Ainsi, que ce soit dans le cadre de sa vie professionnelle ou intime, les relations interpersonnelles qu'un individu entretient avec ses congénères devraient constituer pour lui l'une des sources les plus importantes d'incertitude et de souffrance émotionnelle. Ce faisant, nous faisons l'hypothèse que l'activité cognitive spontanée dans laquelle un individu s'engage la plupart du temps, le *mode par défaut* de sa cognition, consiste en l'exercice de ses habiletés de conceptualisation métacognitive

afin d'élaborer des modèles prédictifs du processus de réalisation de son allostasie à travers les interactions sociales avec ses congénères. Cette activité consiste plus précisément en une tentative de prédire les conséquences des interactions sociales sur son allostasie en modélisant les processus causaux qui en sont responsables : les dynamiques de l'organisation cognitive de l'individu lui-même et des congénères qui composent son environnement social proximal, ainsi que de leur interaction. Cette activité a pour fonction de générer et d'implémenter des inférences actives d'autorégulation émotionnelle et de cognition sociale afin d'optimiser le processus de réalisation de l'allostasie de l'individu en permettant l'implémentation d'un couplage structurel robuste et intime avec les congénères qui composent sa niche écologique. Dans ce contexte, l'association libre peut-être comprise comme une pratique de verbalisation par le client de son activité spontanée de conceptualisation métacognitive, ce qui offre au psychothérapeute la possibilité de l'observer et de collaborer avec lui à l'amélioration des modèles prédictifs qu'elle implémente.

La deuxième technique de la psychothérapie psychodynamique, l'interprétation du transfert, correspond à une pratique dans laquelle le thérapeute observe l'activité comportementale spontanée du client, la lui reflète de manière à ce qu'il puisse en prendre conscience, lui demande de tenter d'expliquer les processus mentaux inconscients qui ont mené à son émergence, et collabore avec lui dans cette tentative d'explication. D'après la compréhension de l'organisation des systèmes neurocognitifs humains qui nous est offerte par le FEP et les approches prédictives, l'activité comportementale spontanée du client devrait correspondre aux inférences actives qui résultent des prédictions générées par les concepts émotionnels qu'il implémente afin d'optimiser le processus de réalisation de son allostasie. Dans ce contexte, le phénomène du transfert peut-être redéfini comme les patterns d'inférence active caractéristiques qui résultent des régularités dans le processus de conceptualisation émotionnelle implicite et automatique à travers lequel un individu réalise son allostasie, des régularités qui proviennent en grande partie du parcours ontogénétique à travers

lequel ce dernier a acquis ses habiletés de conceptualisation métacognitive, un parcours majoritairement déterminé par les relations d'attachement avec ses donneurs de soins primaires. Dans ce contexte, le reflet des comportements du client par le psychothérapeute correspond à la mise en place de régimes d'attention partagée qui dirigent l'attention du client sur les inférences actives qu'il déploie et sur leurs conséquences, et ce afin de renvoyer ce dernier vers l'expérience affective (incertitude ressentie) qui a motivé leur implémentation, et de l'inciter à modéliser le processus causal (émotionnel) qui les lie. Autrement dit, l'interprétation du transfert correspond à une technique à travers laquelle le psychothérapeute cherche à encourager et à influencer le client dans l'exercice de ses habiletés de conceptualisation métacognitive afin de modéliser les processus cognitifs qui sous-tendent son activité comportementale spontanée dans le cadre de la relation thérapeutique.

En résumé, à travers la mise en place de régimes d'attention partagée, l'utilisation des techniques de l'interprétation de l'association libre et du transfert permet au psychothérapeute d'entraîner le client dans un processus collaboratif de reconceptualisation des concepts métacognitifs sous-optimaux qui, en générant des prédictions menant à l'implémentation d'inférences actives d'autorégulation émotionnelle et de cognition sociale non adaptatives, sont responsables de sa souffrance émotionnelle. Les mécanismes spécifiques qui sous-tendent la modification des concepts métacognitifs du client à travers la relation interpersonnelle avec son psychothérapeute peuvent être classés en deux catégories générales qui correspondent à deux types d'apprentissage que nous attribuons respectivement à l'exercice des deux modes des habiletés de conceptualisation métacognitives du client : le mode implicite et automatique, d'une part ; et le mode explicite et contrôlé, d'autre part.

Le premier type d'apprentissage correspond à la modification des concepts métacognitifs du client à travers l'exercice de ses habiletés de conceptualisation métacognitive implicite et automatique durant les interactions récurrentes avec le

psychothérapeute. Ce processus repose sur une modification subtile et indirecte des concepts métacognitifs du client qui résulte de la répétition d'expériences concrètes particulières vécues dans le cadre de son immersion régulière dans le contexte écologique spécifique offert par la psychothérapie psychodynamique, un contexte caractérisé par deux dimensions fondamentales : a) un cadre clinique stable et accueillant, caractérisé par un ensemble de règles prédéfinies et explicites qui déterminent les modalités du traitement (techniques utilisées), la nature de la relation thérapeutique (asymétrique et contractuelle), ainsi que la durée et la fréquence des rencontres (invariantes) ; et b) une relation interpersonnelle avec un donneur de soins ayant suivi une formation professionnelle exigeante afin de développer des habiletés de conceptualisation métacognitive supérieures, ainsi qu'une très bonne maîtrise de son propre fonctionnement psychologique, de manière à ce qu'il soit capable de fournir des soins cognitifs de haute qualité. Les conditions offertes par ce contexte écologique ont pour effet de réduire l'incertitude ressentie par le client vis-à-vis de la pratique de la psychothérapie, et de la relation thérapeutique avec le psychothérapeute. Le cadre clinique stable et accueillant fournit un environnement sécuritaire et confortable pour le client, car il est invariant, et donc *prédictible (certain)*. La capacité du thérapeute à offrir des soins cognitifs efficaces au client, et les attentes claires vis-à-vis de l'attribution de ces soins permise par la nature contractuelle de leur relation, permettent une modulation initiale du prior de confiance épistémique du client suffisante pour enclencher l'installation un couplage structurel suffisamment intime pour permettre une modification significative de son fonctionnement cognitif.

Cependant, le contexte écologique offert par la psychothérapie psychodynamique n'a pas seulement pour effet de permettre ou d'initier l'engagement du client dans cette dernière, mais contribue lui-même au processus d'apprentissage émotionnel qui détermine son efficacité. En effet, l'immersion régulière dans les cycles d'action-perception spécifiques permis par les interactions récurrentes avec un donneur de soins cognitifs compétent dans un contexte environnemental sécuritaire et rassurant va mener

le client à accumuler des expériences relationnelles caractérisées par un faible niveau d'incertitude ressentie (affect positif). Les patterns affectifs, comportementaux, et cognitifs qualitativement différents qui émergent de ces expériences vont mener son activité de conceptualisation métacognitive automatique et implicite (par défaut) à générer de nouveaux concepts implémentant de nouvelles prédictions vis-à-vis de son fonctionnement émotionnel et cognitif, de celui d'autrui, et de leur interaction. Autrement dit, la simple exposition au contexte relationnel positif et sécuritaire offert par la psychothérapie psychodynamique (ou toute autre forme de psychothérapie) devrait mener, à travers un processus d'apprentissage autonome du client, à une modification d'un certain nombre de priors qui conditionnent son fonctionnement émotionnel et relationnel.

Parmi ces priors se trouvent le prior de confiance épistémique, dont la modification initiale devrait être amplifiée à travers l'engagement prolongé dans la psychothérapie, à mesure que le client accumule des expériences de réception de soins cognitifs efficaces. En suivant Fonagy et Allison (2014), nous faisons l'hypothèse que cette modification substantielle du prior de confiance épistémique devrait mener à une ouverture progressive du client vis-à-vis de l'influence relative que l'ensemble des congénères avec qui il entretient des relations interpersonnelles peuvent exercer sur son fonctionnement cognitif. Cette ouverture devrait lui permettre d'enclencher une dynamique exploratoire généralisée dans laquelle il se sert activement des expériences vécues dans ses relations interpersonnelles avec ses congénères afin de s'entraîner à l'exercice habile de ses capacités d'autorégulation émotionnelle et de cognition sociale, et donc d'optimiser le processus de réalisation de son allostasie. Cependant, la capacité du client à choisir adéquatement parmi les donneurs de soins cognitifs potentiels, à déterminer leur influence relative sur son fonctionnement cognitif, et à apprendre efficacement de ses interactions avec ceux-ci va dépendre de l'amélioration de ses habiletés de conceptualisation métacognitive permise par des processus

d'apprentissage qui reposent sur l'influence active du psychothérapeute sur son fonctionnement cognitif.

En effet, le type d'apprentissage émotionnel sous-tendu par l'exercice du mode automatique et implicite des habiletés de conceptualisation métacognitive du client ne se limite pas aux apprentissages autonomes que ce dernier réalise indépendamment de l'influence directe du psychothérapeute. Cette influence, le psychothérapeute l'exerce principalement à l'aide des régimes d'attention partagée qu'il met en place et qu'il dirige. Ces régimes sont caractérisés par une dynamique de direction successive et alternée des ressources attentionnelles du client vers différents sous-ensembles d'indices sensoriels et abstraits qui correspondent aux manifestations comportementales et cognitives des processus causaux qui gouvernent son fonctionnement émotionnel et interpersonnel. Nous pensons que ces régimes peuvent être conçus comme entraînant le client à l'exercice d'un type d'inférences actives particulières, que nous proposons d'appeler des *inférences actives d'échantillonnage informationnel*, qui par le biais du processus d'allocation de précision, opèrent une sélection des informations sur lesquelles va se baser son activité de conceptualisation métacognitive. Ces inférences actives d'échantillonnage informationnel correspondent à des prédictions dont le but est de générer des preuves pour guider le processus d'inférence perceptuelle permettant de la modélisation des processus cognitifs de l'individu.

Selon nous, l'instigation de ces régimes d'échantillonnage informationnel correspond au mécanisme primordial à partir duquel les donneurs de soins primaires induisent l'émergence initiale de l'activité de conceptualisation métacognitive chez les enfants dont ils s'occupent. À travers le mécanisme de l'attention jointe, les donneurs de soins cognitifs poussent un individu à allouer conjointement une plus grande précision à un ensemble d'indices sensoriels et abstraits particuliers, ce qui mène le système prédictif de ce dernier à tenter de modéliser les processus causaux qui les lient. Or, si ces indices

ont été sélectionnés par les donneurs soins, c'est parce que ce sont ceux qu'ils utilisent pour modéliser l'activité cognitive de l'individu : ils sont tous des manifestations des processus qui définissent son organisation neurocognitive. Cela implique que la modélisation prédictive de ces indices devrait mener le système prédictif de l'individu à générer des concepts neurocognitifs qui les expliquent comme résultant d'un même ensemble de causes distales spécifique : sa propre activité cognitive. Autrement dit, l'émergence du processus de conceptualisation métacognitive repose initialement sur la simple direction des ressources attentionnelles d'un individu sur les manifestations observables (extéroceptivement ou intéroceptivement) de ses processus cognitifs. Plus précisément, nous pensons que les inférences actives d'échantillonnage informationnel, en permettant au système prédictif de créer des catégories (*clusters*) d'indices sensoriels et abstraits qui dérivent des processus cognitifs de soi-même ou d'autrui, constituent la fondation du processus de conceptualisation métacognitive.

Les patterns d'inférences actives d'échantillonnage informationnel constituent selon nous les premières prédictions qui définissent un concept métacognitif donné, car elles sont les conditions de possibilité de son implémentation initiale. Par exemple, l'autorégulation émotionnelle requiert d'implémenter un concept émotionnel générant des prédictions perceptuelles et sensorimotrices permettant de minimiser les erreurs de prédictions reliées à un événement émotionnel donné. Or, l'implémentation de ce concept émotionnel requiert nécessairement le déploiement d'inférences actives d'échantillonnage informationnel, qui correspondent à des prédictions dont le but est de générer des preuves pour guider le processus d'inférence perceptuelle permettant de catégoriser l'épisode émotionnel vécu. Plus précisément, le processus de conceptualisation métacognitive qui sous-tend l'autorégulation d'un épisode émotionnel donné correspond à une boucle de rétroaction qui peut être décrite comme suit : tout d'abord, une augmentation dans l'incertitude ressentie (erreurs de prédiction intéroceptives) entraîne l'activation de prédictions sensorimotrices (ou autres) particulières visant à sa minimisation, ce qui correspond à la mise en place d'une

émotion primaire ou basique (Panksepp, 2005, 2006) ; ensuite, des inférences actives d'échantillonnage informationnel spécifiques, sélectionnées à partir des premiers indices sensoriels disponibles, sont implémentées afin de collecter les preuves nécessaires au processus d'inférence perceptuelle permettant la catégorisation adéquate de l'épisode émotionnel en question ; puis cette catégorisation entraîne l'instauration d'un concept émotionnel donné qui implémente des prédictions (inférences actives et perceptuelles) afin de minimiser les erreurs de prédictions expérimentées par le système ; et enfin, l'implémentation de ces prédictions mène à une variation de l'incertitude ressentie, qui entraîne de nouvelles prédictions, et la boucle recommence.

Dans ce contexte, il est logique de considérer que l'activité de conceptualisation métacognitive en général repose un pattern global de direction de l'attention sur les domaines d'indices sensoriels qui sont catégorisés comme des manifestations de l'activité cognitives des êtres humains. Cependant, il est aussi raisonnable de penser que différents concepts métacognitifs demandent des inférences actives d'échantillonnage informationnel particulières afin d'être générés et d'être implémentés. Par exemple, la conceptualisation d'un épisode de peur ou de tristesse ne demande peut-être pas de diriger son attention sur les mêmes domaines d'entrées sensorielles. De plus, la conceptualisation d'états mentaux complexes, comme les intentions d'autrui par exemple, pourrait demander l'implémentation d'inférences actives allant au-delà de simples régimes de direction des ressources attentionnelles, comme des inférences actives de communication langagière par exemple. Tout cela nous mène à penser que les habiletés de conceptualisation métacognitive d'un individu reposent sur un éventail de patterns inférence active d'échantillonnage informationnel particuliers acquis à travers son expérience, et dont les spécificités devraient jouer un rôle important dans la qualité des concepts métacognitifs qu'il développe.

Nous faisons l'hypothèse selon laquelle l'immersion récurrente d'un individu dans les régimes d'attention partagée mis en place par ses donneurs de soins cognitifs devrait

entraîner un apprentissage procédural des patterns d'inférence active d'échantillonnage informationnel spécifiques qu'ils instaurent. Selon nous, ce processus d'apprentissage correspond à la première forme d'influence directe que le psychothérapeute exerce sur les habiletés de conceptualisation métacognitive de son client. Il permettrait à ce dernier d'acquérir des patterns d'inférence active d'échantillonnage informationnel plus efficaces favorisant le développement d'habiletés de conceptualisation métacognitives optimales et adaptatives. Cependant, les régimes d'attention partagée instaurés par le psychothérapeute vont aussi entraîner un autre processus d'apprentissage intrinsèquement lié à celui décrit ci-dessus. Que ce soit dans le contexte de l'association libre ou du transfert, les patterns d'inférence active d'échantillonnage informationnel instaurés par le psychothérapeute vont mener l'activité de conceptualisation métacognitive automatique et implicite du client à se baser sur les mêmes indices sensoriels et abstraits qu'il utilise lui-même pour conceptualiser les processus cognitifs qui gouvernent le fonctionnement émotionnel et interpersonnel de ce dernier. Selon nous, cela devrait mener le client à générer des concepts métacognitifs très similaires à ceux qui gouvernent l'activité de conceptualisation métacognitive du psychothérapeute. Nous proposons d'identifier cette forme d'apprentissage au processus de synchronisation conceptuelle tel que décrit par Gendron et Barrett (2018) : une modification subtile et indirecte des concepts métacognitifs qui gouvernent les capacités d'autorégulation émotionnelle et de cognition sociale du client, basée sur les dynamiques des variations de l'action (comprenant ici les inférences actives d'échantillonnage informationnel) et des changements intéroceptifs (c.-à-d., affectifs) qui ont lieu dans les interactions récurrentes avec son psychothérapeute.

Autrement dit, nous proposons que l'utilisation prolongée de mêmes régimes d'échantillonnage informationnel, causée par le mécanisme universel de l'attention jointe, puisse mener à elle seule à un rapprochement significatif des concepts métacognitifs implémentés par les systèmes prédictifs d'un individu et de son donneur

de soins cognitifs. Dans le cas de la psychothérapie psychodynamique, puisque le psychothérapeute possède des priors plus précis, alors c'est lui qui dirige l'échantillonnage informationnel, ce sont les priors du client qui sont entraînés de manière asymétrique vers les siens (Vasil *et al.*, 2020). Si l'on part du principe que le thérapeute est compétent, et qu'il possède des habiletés de conceptualisation métacognitives supérieures, alors les concepts cognitifs qu'il transmet à son client par le biais du processus de synchronisation conceptuelle devraient être adaptatifs et fonctionnels, et résulter en une amélioration significative de ses capacités d'autorégulation et de cognition sociale. De plus, l'apprentissage parallèle des patterns d'inférence active d'échantillonnage informationnel du thérapeute devrait faire en sorte que la tendance du client à générer des concepts métacognitifs d'un niveau de fonctionnalité similaire à ceux de son psychothérapeute s'étende au-delà du cadre de la relation thérapeutique.

Ici, il est important de rappeler que la mise en place des régimes d'attention partagée qui sous-tendent l'apprentissage des patterns d'inférence active d'échantillonnage informationnel et le processus de synchronisation conceptuelle n'implique pas nécessairement une intention explicite et délibérée de la part du thérapeute (bien que ce soit souvent le cas), et peut tout à fait être réalisée à travers l'exercice du mode automatique et implicite des habiletés de conceptualisation métacognitive de ce dernier. De plus, et c'est la raison pour laquelle nous avons choisi de le présenter ici, nous pensons que le processus de synchronisation conceptuelle s'opère majoritairement (mais peut-être pas totalement) à travers l'exercice du mode implicite et automatique des habiletés de conceptualisation métacognitive du client. En effet, ce processus peut-être conçu comme reposant sur une forme de couplage structurel neurocognitif au sein duquel une congruence structurelle entre deux systèmes prédictifs humains en vient à émerger de manière *bottom-up* à travers leur engagement mutuel et récurrent dans des cycles d'action-perception collaboratifs. Cependant, le couplage structurel neurocognitif qui s'installe entre un client et son psychothérapeute dans le

cadre de la psychothérapie psychodynamique s'opère aussi selon un autre processus que l'on pourrait qualifier de *top-top*, un processus qui repose principalement sur l'exercice collaboratif du mode explicite et contrôlé de leurs habiletés de conceptualisation métacognitive. C'est ce processus que la plupart des tentatives préalables d'explication de la psychothérapie psychodynamique, comme le modèle proposé par Solms (2018) par exemple, identifient comme le mécanisme principal du changement thérapeutique sous-tendu par cette pratique, et c'est vers lui que nous allons nous tourner maintenant.

Afin d'expliciter ce mécanisme, il nous faut rendre compte de l'un des objectifs principaux de l'utilisation des techniques thérapeutiques de l'interprétation de l'association libre et du transfert que nous avons intentionnellement éludé jusqu'à présent. En effet, en encourageant le client à diriger activement ses ressources attentionnelles ainsi que l'ensemble de son activité cognitive vers la modélisation des processus cognitifs qui sous-tendent son fonctionnement émotionnel et interpersonnel spontané, ces techniques ont très souvent pour conséquence d'éliciter chez ce dernier l'exercice du mode explicite et contrôlé de ses habiletés de conceptualisation métacognitive.

Comme nous l'avons proposé précédemment, l'association libre consisterait en une pratique communicationnelle de verbalisation par le client de son activité spontanée de conceptualisation métacognitive de ses expériences émotionnelles et interpersonnelles passées ou présentes. Cette pratique implique l'engagement du client dans une activité d'autorégulation émotionnelle et de cognition sociale basée sur la manipulation de représentations langagières et mnésiques afin de guider la conceptualisation des processus cognitifs qui sont responsables de la souffrance émotionnelle qui émerge de sa vie personnelle, professionnelle, etc. L'interprétation du transfert consisterait quant-à-elle à une pratique visant à entraîner le client dans une activité de conceptualisation métacognitive des processus cognitifs qui ont mené à l'émergence de l'activité

comportementale spontanée qu'il manifeste dans le cadre de la psychothérapie. De même que dans le cas de l'association libre, le client doit verbaliser son activité de conceptualisation métacognitive pour pouvoir la communiquer au psychothérapeute. En résumé ces deux techniques entraînent le client dans une activité de conceptualisation métacognitive de ses propres processus cognitifs médiatisée par l'utilisation du langage.

Ici, il est primordial de prendre conscience d'un état de fait crucial : les processus cognitifs que le client est incité à modéliser à travers l'utilisation de ces deux techniques thérapeutiques correspondent en grande partie aux processus qui sous-tendent la grande majorité de son fonctionnement émotionnel et interpersonnel, c'est-à-dire, à l'exercice du mode implicite et automatique de ses habiletés de conceptualisation métacognitive. Autrement dit, l'interprétation du transfert et de l'association libre renvoie le client au processus de conceptualisation métacognitive automatique qui conditionne son fonctionnement émotionnel et interpersonnel actuel et passé, et réalise cela en dirigeant son attention vers, c'est-à-dire, en explicitant, ses manifestations cognitives, affectives et comportementales implicites, qu'il expérimentait auparavant sur le mode de la conscience préreflexive. En incitant le client à s'engager dans une forme de cognition représentationnelle dans laquelle il utilise le langage afin de réaliser une prise de conscience réflexive de son activité de conceptualisation métacognitive spontanée, l'objectif principal de ces techniques est de l'aider à exercer un contrôle exécutif sur cette dernière, ce qui revient à la faire passer d'un mode automatique et implicite vers un mode explicite et contrôlé.

Nous faisons l'hypothèse que l'engagement récurrent dans l'interprétation de l'association libre et du transfert devrait entraîner un processus d'apprentissage permettant une atténuation progressive de la prédominance excessive de la partie automatique et implicite des processus de conceptualisation métacognitive du client, que nous avons associé précédemment à la génération et à l'implémentation de

concepts métacognitifs sous-optimaux entraînant des inférences actives d'autorégulation émotionnelle et de cognition sociale non adaptatives. Tout d'abord, la présence rassurante et régulatrice d'un donneur de soins cognitif compétent devrait permettre au client de tolérer l'élévation temporaire de l'incertitude ressentie liée à son engagement dans le mode de cognition exploratoire et relativement risqué soutenu par l'exercice de ses habiletés de conceptualisation métacognitive explicite et contrôlée (REF). La répétition d'expériences positives (c-à-d., présentant un faible niveau d'incertitude ressentie) reliées à l'utilisation efficace de ce mode de métacognition va fournir au système prédictif du client un ensemble de preuves de son potentiel pour la minimisation de son énergie libre. De plus, ces expériences vont fournir au client un ensemble de souvenirs épisodiques qui vont pouvoir servir d'outils mnésiques et représentationnels essentiels au type de cognition déclarative nécessaire à l'exercice du mode explicite et contrôlé de ses habiletés de conceptualisation métacognitives. En effet, nous avons proposé dans notre section sur l'inconscient dynamique que l'une des raisons principales de la dominance exagérée des concepts métacognitifs implicites corresponde à l'absence relative de souvenirs épisodiques reliés aux expériences qui ont mené à leur génération, privant le système prédictif de représentations explicites à travers lesquelles il pourrait opérer leur prise de conscience réflexive, l'empêchant ainsi de les rappeler et de les manipuler dans la mémoire de travail afin de les soumettre à un contrôle exécutif permettant leur reconceptualisation (reconsolidation ; Solms, 2018).

Povinelli et Simon (1998) ont proposé que ce soit l'amélioration des capacités de cognition représentationnelles qui mènerait à l'émergence de la cognition autobiographique vers l'âge de 4 ans. Selon eux, c'est la capacité de maintenir plusieurs souvenirs et concepts reliés au soi dans la mémoire de travail, et à établir des relations temporelles et causales entre elles, qui permettrait la constitution du concept de *soi autobiographique* (Fonagy *et al.*, 2018). Dans la recherche, l'émergence de la cognition autobiographique est systématiquement associée au développement de la modalité

narrative de l'expression langagière, de telle manière que les concepts de soi autobiographique et de soi narratif soient considéré comme fondamentalement interreliés, sinon équivalents (Fivush, 2011 ; Fivush et Haden, 2003 ; Singer et Blagov, 2004 ; Singer et Bluck, 2001). En effet, le langage narratif devrait constituer un outil central dans le processus de conceptualisation autobiographique, en permettant la combinaison des concepts afin d'élaborer un modèle dynamique de l'évolution des états mentaux de l'individu à travers le temps. Tel que nous l'avons vu précédemment, la mentalisation (Denny *et al.*, 2012 ; Frith *et al.*, 2003 ; Hyatt *et al.*, 2015), la mémoire autobiographique (Hoffmann, 2020 ; Spreng *et al.*, 2008), et la cognition autoréférentielle reliée au concept de soi (Molnar-Szakacs et Uddin, 2013 ; Moran *et al.*, 2013) ont toutes été associées aux mêmes substrats neuronaux, à savoir les nœuds du réseau du mode par défaut. De plus, l'activité cognitive à laquelle ce dernier a été associé à l'origine, le type de cognition dans lequel nous nous trouvons tous lorsque nous sommes dans un état de repos non contraint, est consistante avec le processus de conceptualisation qui, tel que décrit ci-dessus, serait à l'origine du soi autobiographique ou narratif. En effet, cette activité mentale correspond à une pensée autoréférentielle relativement détachée du moment présent, et élaborant des modèles prospectifs du futur à l'aide informations issues des expériences passées (Carhart-Harris et Friston, 2010 ; Raichle, 2015 ; Raichle *et al.*, 2001).

Nous proposons que la conceptualisation autobiographique puisse être conçue, au moins en partie, comme une forme de conceptualisation métacognitive explicite et contrôlée d'un niveau d'abstraction temporelle élevée, que nous identifions à un processus de modélisation dynamique de l'évolution du fonctionnement cognitif d'un individu à travers l'histoire des interactions récurrentes avec sa niche écologique. Plus précisément, nous pensons que cette forme explicite de conceptualisation autobiographique pourrait avoir pour fonction de suivre l'évolution des concepts métacognitifs développés et employés par l'individu à travers le temps, et ce afin d'évaluer leur efficacité relative dans le processus de réalisation de son allostasie et

d'instauration d'un couplage structurel robuste avec sa niche écologique. Selon nous, l'utilisation de la modalité narrative du langage, afin d'élaborer un modèle dynamique de l'évolution de ses concepts métacognitifs et de leur efficacité dans la minimisation de son énergie libre, constitue l'une des manières principales à travers laquelle un système prédictif humain peut guider leur reconceptualisation optimale. Ce processus correspond à la modélisation prédictive des causes distales du développement de ces concepts, qui correspondent aux propriétés des situations concrètes qui ont mené à leur émergence. Ces situations correspondent à des agencements particuliers du système cerveau-corps-environnement, des agencements qui sont grandement déterminés par les dynamiques des interactions sociales récurrentes entre un être humain et ses congénères qui sont nécessaires pour la réalisation de son allostase.

Nous faisons l'hypothèse que le caractère sous-optimal de certains concepts métacognitifs repose dans le fait qu'ils encodent des modèles prédictifs reliés à des situations provenant du passé développemental de l'individu, des modèles qui ne sont pas adaptatifs dans le contexte écologique dans lequel il évolue aujourd'hui. Bien souvent, ces modèles peuvent ne jamais avoir été particulièrement efficaces, mais constituaient plutôt des solutions imparfaites à des problèmes insolubles reliés à des situations environnementales défavorables (ex. donneurs de soins incompetents ; Solms, 2018). D'après nous, le fait que ces modèles continuent d'être implémentés dans des situations différentes repose sur la dynamique d'organisation neurocognitive pathogène que nous avons associé à la dominance exagérée du mode implicite et automatique du processus de conceptualisation métacognitive, à savoir, la tendance à favoriser l'évitement proximal de l'augmentation temporaire de l'incertitude ressentie (intensité affective) à un niveau qui est prédit comme étant intolérable pour le système. En effet, le caractère défavorable des situations mentionnées ci-dessus est défini par le fait qu'elles entraînent une incapacité relative du système prédictif de l'individu à minimiser son énergie libre efficacement, ce qui correspond à l'expérimentation d'un haut niveau d'incertitude ressentie vis-à-vis du processus de réalisation de son

allostasie. Ainsi, les concepts métacognitifs qui émergent de l'immersion dans ces situations encodent des attentes (prior beliefs) qui attribuent un haut niveau d'incertitude ressentie vis-à-vis des conditions du système cerveau-corps-environnement qui les définissent. Nous faisons l'hypothèse que la simple exposition à certaines de ces conditions, caractérisée par la survenue d'ensembles d'erreurs de prédictions particuliers, pourrait mener à l'émergence d'un niveau d'incertitude ressentie suffisant pour donner lieu à l'inhibition du mode explicite et contrôlé du processus de conceptualisation métacognitive au profit de son mode implicite et automatique, et ce dans le but d'éviter l'augmentation intolérable de l'énergie libre qui est anticipée par le système prédictif.

De plus, nous proposons que l'inhibition du mode implicite et contrôlé de ses habiletés de conceptualisation métacognitive puisse empêcher l'individu d'intégrer à ses concepts métacognitifs les souvenirs épisodiques qui encodent les situations concrètes qui ont mené à leur émergence, ainsi que celles qui mènent à leur implémentation continue. Plus précisément, ce n'est pas que ces souvenirs épisodiques sont nécessairement absents, mais que les expériences qu'ils encodent n'ont pas été explicitement associées aux concepts qu'elles ont menés à développer. Ainsi, nous pensons que la prédominance excessive du mode implicite et automatique des habiletés de conceptualisation métacognitive d'un individu peut mener à un appauvrissement des ressources mnésiques nécessaires à l'exercice de ses habiletés de conceptualisation autobiographique. Sans ces ressources, l'individu se retrouverait incapable de procéder à la modélisation efficace de l'évolution et de l'efficacité de ses concepts métacognitifs qui serait nécessaire pour guider leur reconceptualisation optimale. Selon nous, cela devrait mener à l'allocation d'un faible niveau de précision alloué aux prédictions générées durant les tentatives de reconceptualisation explicite et contrôlée des concepts métacognitifs sous-optimaux, et donc à une assise supplémentaire de la domination du mode implicite et contrôlé des processus de conceptualisation métacognitive.

L'utilisation des techniques de l'interprétation de l'association libre et du transfert, en sollicitant chez le client l'exercice du mode explicite et contrôlé de ses habiletés de conceptualisation métacognitive, devrait selon nous entraîner une réactivation progressive du processus de conceptualisation autobiographique permettant la reconsolidation de ses concepts métacognitifs sous-optimaux. Tout d'abord, l'un des objectifs de l'association libre est de mener le patient à se remémorer ses expériences passées, lointaines ou récentes, qui sont les plus pertinentes pour comprendre d'où provient sa souffrance émotionnelle. Ces expériences peuvent-êtres classées en deux catégories : a) les expériences qui ont mené à l'émergence de ses concepts métacognitifs sous-optimaux ; et b) les expériences au sein desquelles il implémente ses concepts métacognitifs sous-optimaux. L'interprétation de l'association libre correspond à une pratique collaborative de conceptualisation métacognitive explicite et contrôlée des situations reliées à ces expériences passées, de manière à : 1) identifier les concepts métacognitifs qui les sous-tendent, et associer les souvenirs épisodiques qui les encodent à ces derniers ; 2) entamer un processus de conceptualisation autobiographique permettant d'identifier les conditions qui ont mené à leur émergence et qui sous-tendent leur implémentation continue, ainsi que de réaliser une prise de conscience réflexive de leur caractère sous-optimal ou pathogène ; et 3) procéder à leur reconceptualisation optimale afin qu'ils implémentent de nouvelles inférences actives d'autorégulation émotionnelle et de cognition sociale plus fonctionnelles et adaptatives.

Cependant, il est important de noter que les concepts métacognitifs sous-optimaux proviennent souvent d'expériences ontogénétiques précoces pour lesquelles l'individu ne possède aucun souvenir épisodique. De plus, l'activité cognitive spontanée du client qui se manifeste dans l'association libre peut souvent être influencée par la tendance à l'évitement proximal de l'augmentation temporaire de l'incertitude ressentie (refoulement), ce qui peut mener celui-ci à éviter de se remémorer les expériences affectives intenses reliées à ses concepts métacognitifs sous-optimaux. C'est pour cette raison que la technique de l'association libre est combinée à celle de l'interprétation du

transfert, qui se concentre sur la conceptualisation du fonctionnement cognitivo-comportemental démontré par le client dans le cadre de la relation avec son psychothérapeute. L'objectif de cette technique est toujours d'identifier et de reconceptualiser les concepts métacognitifs sous-optimaux du client, mais les expériences vécues à travers son utilisation vont permettre d'entraîner le client à l'exercice en temps réel de ses habiletés de conceptualisation métacognitive explicite et contrôlée et des capacités d'autorégulation émotionnelle et de cognition sociale qu'elles sous-tendent.

Premièrement, cet entraînement va permettre l'acquisition et l'accumulation traces mnésiques explicites qui vont pouvoir servir de support représentationnel afin de soutenir la reconceptualisation explicite des concepts métacognitifs sous-optimaux en permettant d'exercer un contrôle exécutif sur le processus implicite de conceptualisation métacognitive automatique qui sous-tend leur implémentation. En effet, nous faisons l'hypothèse que l'engagement dans l'exercice du mode explicite et contrôlé des habiletés de conceptualisation métacognitive repose sur lui aussi sur l'implémentation de patterns d'inférences actives d'échantillonnage informationnel particuliers, et que l'exercice de l'interprétation du transfert permet leur apprentissage. Cependant, ces inférences actives ne consistent pas simplement à diriger les ressources attentionnelles vers certains ensembles d'entrées sensorielles ou d'erreurs de prédiction, mais aussi à se servir des indices fournis par ces derniers pour aller puiser dans la mémoire afin de trouver des informations permettant de renseigner et de guider l'activité de conceptualisation métacognitive de l'événement neurocognitif vécu. Les informations ainsi puisées dans la mémoire correspondent selon nous à des représentations qui, lors des expériences passées, ont été utilisées dans le processus de conceptualisation explicite et contrôlée de certains événements cognitifs, de telle manière qu'elles ont été intégrées aux concepts métacognitifs dont elles ont participé à l'élaboration, et que leur rappel puisse permettre la réinstanciation de ces concepts. Ces représentations ne comportent pas seulement à des images extéroceptives

correspondant aux souvenirs épisodiques reliés aux expériences passées, mais aussi et surtout à des mots, des symboles langagiers utilisés pour référer aux événements cognitifs, ainsi que pour les catégoriser et les expliquer, c'est-à-dire, pour les *conceptualiser*.

Selon nous, c'est à travers l'introduction de représentations langagières dans l'activité de conceptualisation métacognitive des états émotionnels des enfants dont ils s'occupent que les donneurs de soins primaires instiguent l'émergence du mode explicite et contrôlé des habiletés de conceptualisation métacognitive de ces derniers. En effet, lors de l'activité collaborative de conceptualisation émotionnelle implicite instaurée par les régimes d'attention partagée qu'ils mettent en place, les donneurs de soins primaires introduisent des stimuli auditifs saillants, des mots, qui à force d'être systématiquement associés à des concepts émotionnels, vont être intégrés aux modèles prédictifs qui leur correspondent. Leur forme déclarative leur permettant d'être facilement rappelés à la mémoire de travail, ces représentations extéroceptives vont servir de support mnésique facilitant l'instanciation des concepts neurocognitifs auxquels ils sont associés (Gendron et Barrett, 2018 ; Solms, 2018). À partir du moment où l'enfant aura fermement associé un mot à un concept émotionnel donné, sa simple prononciation par ses donneurs de soins devrait avoir le potentiel d'éliciter de manière relativement directe l'instanciation de ce dernier, et entraîner la cascade de prédictions pertinentes à la régulation de l'épisode émotionnel correspondant. Cette instanciation d'un concept neurocognitif à travers la réception d'un stimulus vocal auquel il est associé par une relation de représentation symbolique constitue selon nous l'une des premières instances de communication langagière. Atzil *et al.* (2018) ont proposé que la motivation intrinsèque (cf. réalisation de l'allostasie) reliée aux interactions de modulation homéostatique et affective qui s'opèrent entre un enfant et ses donneurs de soins sous-tende l'apprentissage des habiletés cognitives complexes qui s'opère durant celles-ci. En suivant cette proposition, nous faisons l'hypothèse que l'un des motivateurs centraux de l'apprentissage du langage consisterait en son utilité pour

l'implémentation des concepts émotionnels qui permettraient la régulation affective de l'enfant.

À partir du moment où il aura développé des capacités de contrôle exécutif suffisantes, l'enfant devrait devenir capable de reconnaître certains indices sensoriels typiques de catégories d'états émotionnels particulières, de se souvenir des mots et/ou des images extéroceptives qui représentent ces catégories, et donc de procéder de manière autonome à l'instanciation des concepts neurocognitifs permettant de réguler son état émotionnel. Dans le cadre des approches prédictives, cette activité de conceptualisation implique une suppression des erreurs de prédictions intéroceptives et sensorimotrices en implémentant des prédictions descendantes qui les expliquent et réduisent l'incertitude ressentie par le système prédictif de l'enfant. En donnant lieu à une régulation autonome de l'expérience affective et des comportements de l'enfant, le processus décrit ci-dessus peut selon nous être identifié à l'exercice de ce que l'on appelle communément les capacités d'autorégulation émotionnelle. Ce processus de génération endogène d'images extéroceptives (auditives ou visuelles) ayant pour fonction de référer à, et d'éliciter l'instanciation d'un concept neurocognitif donné, correspondrait selon nous à l'une des premières instances de cognition représentationnelle, qui comme nous le savons, va constituer l'une des dimensions les plus importantes de la cognition humaine. En effet, ce type de cognition sous-tend la capacité de penser, qui correspond à l'activité cognitive dans laquelle nous nous trouvons la plupart du temps : notre mode par défaut.

Nous proposons que le processus de renforcement des habiletés de conceptualisation métacognitive explicite et contrôlée du client qui s'opère dans le cadre de la psychothérapie psychodynamique repose tout d'abord sur le même processus qui avait permis leur émergence dans le cadre de sa relation avec ses donneurs de soins primaires. Dans l'interprétation du transfert, le psychothérapeute participe activement au processus de conceptualisation métacognitive de l'activité cognitivo-comportementale

spontanée du client, et l'une de ses contributions majeures consiste introduire des inférences actives de communications langagières, des *reflets*, qui réfèrent aux manifestations de cette activité et qui interprètent explicitement ces dernières comme des indices des processus cognitifs qui les sous-tendent. En plus de constamment solliciter le patient à l'exercice du mode explicite et contrôlé de ces habiletés de conceptualisation métacognitive, ces interventions du psychothérapeute ont pour effet de fournir un nouvel ensemble de mots et d'expressions langagières qui vont pouvoir être utilisés par le client dans cet exercice, et qui vont être intégrés aux concepts métacognitifs qu'elle permet de générer. En plus de servir au processus de conceptualisation en tant que tel, ces représentations vont constituer un arsenal d'outils mnésiques à partir desquels le client va pouvoir déclencher l'exercice du mode explicite et contrôlé de ses habiletés de conceptualisation métacognitive.

Comme proposé précédemment, nous pensons que les cycles d'action-perception collaboratifs instaurés par l'interprétation du transfert ont pour conséquence première d'entraîner le client à l'exercice de patterns d'inférence active d'échantillonnage informationnel qui visent spécifiquement à l'implémentation du mode explicite et contrôlé de ses habiletés de conceptualisation métacognitives. Ces patterns consistent en une direction initiale des ressources attentionnelles sur des ensembles particuliers d'indices sensoriels ou abstraits, puis à l'engagement immédiat dans une forme de cognition représentationnelle basée sur le rappel de mots ou d'expressions langagières préalablement utilisées pour décrire et catégoriser ces indices. L'engagement dans cette forme de cognition représentationnelle permettrait une prise de conscience réflexive de, et l'exercice d'un contrôle exécutif sur, l'activité de conceptualisation métacognitive automatique et implicite dont les indices détectés constituent des manifestations. Cet exercice permettrait l'instanciation de concepts métacognitifs explicites générant des prédictions qui vont venir minimiser les erreurs de prédictions liées à l'instanciation préalable et automatique des concepts métacognitifs sous-optimaux implémentés par la mentalisation implicite.

L'apprentissage des patterns d'inférence active d'échantillonnage informationnel décrits ci-dessus devrait selon nous mener à une augmentation significative de la tendance du client à s'engager spontanément dans l'exercice du mode explicite et contrôlé de ses habiletés de conceptualisation métacognitive, et donc à une atténuation de la dominance exagérée de son mode implicite et automatique qui serait à l'origine des ses concepts métacognitifs sous-optimaux. De plus, l'alternance dynamique entre l'interprétation du transfert et de l'association libre va permettre au client d'associer ses représentations langagières nouvellement acquises, ainsi que les concepts métacognitifs explicites auxquels elles correspondent, à d'autres indices sensoriels et abstraits correspondant à des expériences passées vécues en dehors de la thérapie. Cela devrait avoir pour conséquence de permettre au client de lier ces représentations aux conditions qui caractérisent ces expériences, de manière à solliciter leur rappel, et donc le déclenchement du mode explicite et contrôlé de ses habiletés de conceptualisation métacognitive, lorsqu'il sera confronté de nouveau à des situations similaires.

Nous pouvons maintenant passer au dernier aspect du processus de modification de l'organisation neurocognitive du client qui s'opère à travers le couplage structurel avec son psychothérapeute dans le cadre de la psychothérapie psychodynamique. Ce processus correspond à la deuxième forme d'apprentissage sous-tendue par l'exercice collaboratif du mode explicite et contrôlé des habiletés de conceptualisation métacognitive du client et de son psychothérapeute, et repose sur une pratique de communication langagière à travers laquelle le psychothérapeute partage directement et explicitement à son client sa propre compréhension de son fonctionnement cognitif. Ce processus est celui qui est le plus souvent mis de l'avant dans les tentatives d'explication de la psychothérapie psychodynamique, ce qui selon nous est dû au fait qu'il s'agit du processus le plus évident et explicite. De plus, il correspond à la forme principale des soins qu'un individu peut recevoir de la part de l'ensemble de ses donneurs de soins cognitifs potentiels, des soins qui jouent un rôle crucial dans l'apprentissage des formes complexes de cognition qui sous-tendent le développement

de capacités d'autorégulation émotionnelle et de cognition sociale adaptative et fonctionnelle.

Vers l'âge de 3 ou 4 ans, l'amélioration des habiletés langagières et des fonctions exécutives d'un enfant humain va l'émergence d'une capacité à maintenir et à manipuler plusieurs concepts dans sa mémoire de travail (Fonagy *et al.*, 2018 ; Povinelli, 1995). Cela va permettre à ses donneurs de soins de lui communiquer une conception de plus en plus complexe et détaillée du processus causal qui conditionne ses états mentaux. En effet, l'utilisation du langage ne s'arrête bien évidemment pas à l'attribution de symboles linguistiques uniques à des concepts émotionnels ou mentaux individuels. Différents concepts générés à partir d'expériences distinctes associées à diverses situations vont pouvoir être combinés et modifiés par une activité mentale autonome qui va devenir de moins en moins dépendante de nouvelles expériences concrètes. Selon nous, ce processus de modification rétrospective de concepts neurocognitifs élaborés à partir d'expériences passées correspond directement à ce que Solms (2013, 2018) appelle la reconsolidation neuronale. C'est à partir de ce moment-là que peut émerger le type de communication langagière mentionné dans la description que Gendron et Barrett (2018) donnent du phénomène de *co-construction conceptuelle*. Dans ce contexte, les donneurs de soins cognitifs peuvent, en communiquant à l'enfant leur conceptualisation de ses processus cognitifs, générer chez lui de tout nouveaux concepts qui vont instancier de toutes nouvelles prédictions permettant d'expliquer et de réguler son activité cognitive d'une toute nouvelle manière.

Au fur et à mesure du développement cognitif et langagier de l'enfant, l'activité de modélisation prédictive de son fonctionnement affectif, cognitif et comportemental en relation à son environnement devrait mener à l'élaboration de concepts de plus en plus abstraits et couvrant des échelles spatiotemporelles et informationnelles de plus en plus étendues. À l'image de l'organisation du cerveau, ces concepts vont former des échafaudages hiérarchiques au sein desquels l'activité de conceptualisation des niveaux

supérieurs porte sur l'information issue de l'activité de conceptualisation des niveaux inférieurs. En haut de cette hiérarchie se trouvent des concepts neurocognitifs comme la personnalité ou l'identité, qui expliquent et régulent l'activité cognitive de l'individu en la contextualisant vis-à-vis de la place qu'il occupe dans son écosystème socioculturelle. Si nous reformulons la position de Carhart-Harris et Friston (2010) présentées au début de ce chapitre à partir du travail réalisé jusqu'ici, tout en haut de cette hiérarchie devrait se trouver le concept d'Ego ou de Soi : le modèle le plus abstrait de l'organisme et de sa situation dans son monde généré par le cerveau humain. Ce concept neurocognitif serait issu de l'intégration et de l'implémentation dynamique de l'ensemble des modèles prédictifs du soi à différentes échelles spatiotemporelles et informationnelles de l'organisation d'un système neurocognitif humain et du couplage structurel avec sa niche écologique : allant du soi minimal corporel et affectif jusqu'au soi autobiographique et narratif, en passant par le soi psychologique.

Dans ce contexte, nous faisons l'hypothèse que la forme principale des processus d'apprentissage de nature *top-top* qui sont responsables de la modification des concepts métacognitifs du client dans le cadre de la psychothérapie repose sur une pratique de communication langagière au sein de laquelle ce dernier et son psychothérapeute s'engagent dans une activité de co-construction conceptuelle telle que décrite par Gendron et Barrett (2018). Dans cette pratique, le client et son psychothérapeute utilisent des inférences actives de communication langagière afin de mettre en place un processus de co-conceptualisation métacognitive explicite et contrôlée des processus cognitifs du client à partir de ses manifestations actuelles (transfert) et passées (association libre). L'utilisation des formes complexes du langage va leur permettre de développer des modèles explicatifs précis des processus causaux qui sont responsables du fonctionnement émotionnel et interpersonnel du client, des modèles pouvant couvrir différentes échelles temporelles, allant de la simple réaction émotionnelle reliée à une interaction sociale jusqu'aux dynamiques de l'évolution du fonctionnement cognitif de l'individu à travers les années et les relations interpersonnelles.

Au sein de cette pratique de co-construction conceptuelle, le psychothérapeute peut, en communiquant à l'individu sa conceptualisation de ses processus cognitifs, générer chez lui de tout nouveaux concepts métacognitifs qui vont instancier de toutes nouvelles prédictions permettant d'expliquer et de réguler son activité cognitive d'une toute nouvelle manière. Par exemple, il peut directement procéder à la conceptualisation métacognitive explicite et contrôlée de l'activité de conceptualisation métacognitive automatique et implicite du client à partir de ses manifestations actuelles dans le transfert. En explicitant leurs manifestations, le psychothérapeute peut entraîner chez le client la prise de conscience réflexive des inférences actives d'autorégulation émotionnelle et de cognition sociale non adaptatives qui résultent de prédictions issues de ses concepts métacognitifs sous-optimaux. Il peut procéder pour lui à la description des priors encodés par ces concepts métacognitifs, et lui démontrer la manière optimale de reconceptualiser ces derniers. Plus particulièrement, il peut lui montrer des manières alternatives de procéder à l'inférence perceptuelle des processus cognitifs qui sous-tendent son comportement ou celui d'autrui, et lui proposer des inférences actives d'autorégulation émotionnelles et de cognition sociale différentes et plus fonctionnelles que celles qu'il avait déployées de lui-même.

Nous faisons l'hypothèse que l'explicitation, la mise en commun, et la combinaison des conceptualisations du client et du psychothérapeute devraient mener au développement d'un modèle commun du fonctionnement cognitif du client. D'après le modèle de la communication coopérative proposé par Vasil *et al.* (2020), puisque le psychothérapeute possède des priors plus précis, ce sont les priors du client qui sont entraînés de manière asymétrique vers les siens. Si l'on part du principe que le psychothérapeute est compétent, et qu'il possède des habiletés de conceptualisation métacognitives supérieures, alors les concepts métacognitifs qu'il transmet à son client par le biais du processus de co-construction conceptuelle devraient être adaptatifs et

fonctionnels, et résulter en une amélioration significative de ses capacités d'autorégulation émotionnelle et de cognition sociale.

En résumé, nous avons proposé que l'amélioration des habiletés de conceptualisation métacognitives du client à travers l'exercice collaboratif de leur mode explicite et contrôlé avec le psychothérapeute repose sur deux mécanismes principaux : a) l'apprentissage de patterns d'inférence active d'échantillonnage informationnel qui permettent favoriser le déclenchement du mode explicite et contrôlé de ses habiletés de conceptualisation métacognitive, et d'atténuer la domination de leur mode implicite et automatique ; b) l'identification et la reconceptualisation des concepts métacognitifs sous-optimaux qui sous-tendent les inférences actives d'autorégulation émotionnelle et de cognition sociale non adaptatives qui sont responsables de sa souffrance émotionnelle ; et c) l'acquisition de nouveaux concepts métacognitifs plus optimaux lui permettant de mettre en place des inférences actives d'autorégulation émotionnelle et de cognition sociale adaptatives et fonctionnelles lui permettant d'optimiser efficacement le processus de réalisation de son allostasie tout en instaurant un couplage structurel robuste avec son éconiche socioculturelle. Nous faisons l'hypothèse que cette amélioration se combine de manière synergique avec celle qui résulte de l'influence du psychothérapeute sur le mode implicite et automatique des habiletés de conceptualisation métacognitives du client, une influence qui repose sur : a) l'apprentissage procédural des patterns d'inférence active d'échantillonnage informationnel optimaux permettant à l'activité de conceptualisation métacognitive automatique et implicite de générer et d'implémenter des concepts métacognitifs adaptatifs ; et b) le processus bottom-up de synchronisation conceptuelle du client avec son psychothérapeute, basé sur l'utilisation de mêmes patterns d'échantillonnage informationnel.

Enfin, l'amélioration globale des habiletés de conceptualisation métacognitive du client devrait lui permettre d'entamer un processus d'apprentissage autonome à partir

des expériences vécues dans ses relations interpersonnelles avec d'autres donneurs de soins cognitifs en dehors de la psychothérapie. En effet, nous avons proposé précédemment que la relation thérapeutique avec un donneur de soins cognitifs compétent offerte par la psychothérapie psychodynamique puisse mener à une modification substantielle du prior de confiance épistémique du client (Fonagy et Allison, 2014). Cette modification mènerait à une ouverture progressive du client vis-à-vis de l'influence relative que l'ensemble des congénères avec qui il entretient des relations interpersonnelles peuvent exercer sur son fonctionnement cognitif. Cette ouverture devrait lui permettre d'enclencher une dynamique exploratoire généralisée dans laquelle il se sert activement des expériences vécues dans ses relations interpersonnelles avec ses congénères afin de s'entraîner à l'exercice habile de ses capacités d'autorégulation émotionnelle et de cognition sociale. Nous faisons l'hypothèse selon laquelle l'amélioration globale des habiletés de conceptualisation métacognitive du client dans le cadre de la psychothérapie psychodynamique devrait lui permettre de choisir adéquatement parmi les donneurs de soins cognitifs potentiels qui s'offrent à lui, de déterminer l'influence optimale qu'il va leur permettre d'exercer sur son fonctionnement cognitif, et d'apprendre efficacement de ses interactions avec ceux-ci.

Cette généralisation du processus d'apprentissage permettant l'amélioration des habiletés de conceptualisation métacognitive du client aux relations interpersonnelles entretenues en dehors de la psychothérapie constitue selon nous un marqueur de l'amélioration substantielle des capacités d'autorégulation émotionnelle et de cognition sociale du client. Cette amélioration devrait coïncider avec une atténuation significative de sa souffrance émotionnelle, c'est-à-dire, avec l'établissement d'une dynamique de métastabilité permettant la baisse globale de l'énergie libre du système vivant qu'il incarne, car elle devrait permettre d'optimiser le processus de réalisation de son allostase en instaurant un couplage cognitif robuste avec sa niche écologique.

CONCLUSION

Dans ce mémoire de maîtrise, nous avons tenté de défendre la thèse suivante :

- Il est possible de formuler une explication des phénomènes en jeu au sein de la psychothérapie psychodynamique qui soit compatible avec les intuitions fondamentales de la psychanalyse freudienne et de la psychologie psychodynamique, tout en étant réductible à des théories scientifiques actuelles, que ce soit en psychologie, en neurosciences cognitives, en anthropologie cognitive, etc.

Afin de défendre cette thèse, ce mémoire a constitué une tentative d'explication naturaliste et mécaniste de la pratique de la psychothérapie psychodynamique à l'aide de modèles théoriques prometteurs issus des avancées récentes réalisées au sein des neurosciences cognitives contemporaines : le principe de l'énergie libre et son extension dans les approches prédictives du cerveau humain. Pour ce faire, notre mémoire s'est divisé en trois étapes.

Dans la première partie de notre mémoire intitulée *explanandum* (ce qui est à expliquer), nous avons procédé à une introduction à la théorie psychanalytique et à la psychologie psychodynamique contemporaine, afin de nous constituer une compréhension initiale des phénomènes sur lesquels reposent la pratique de la psychothérapie psychodynamique, ainsi qu'une conception provisoire des mécanismes dont dépendent son fonctionnement et son efficacité. Au sein de la psychologie psychodynamique, nous avons choisi deux modèles théoriques contemporains, la Neuropsychanalyse et la

théorie de la mentalisation, qui nous ont permis de formuler un ensemble d'hypothèses qui sous-tendent la pratique de la psychothérapie psychodynamique dans un langage conceptuel propice à leur reformulation ultérieure dans les termes des théories qui composent les neurosciences cognitives contemporaines.

Le travail réalisé dans cette première partie nous a permis de formuler une définition initiale de la psychothérapie psychodynamique comme d'une forme d'intervention psychosociale qui consiste à mettre en place une relation interpersonnelle aux qualités spécifiques (ex. attachement sécuritaire) entre un client et un professionnel entraîné (psychothérapeute), et ce afin de recréer les conditions d'acquisition du fonctionnement émotionnel et relationnel du patient dans le but de le modifier. Cette modification passerait par deux mécanismes principaux : a) la prise de conscience, par le patient, de son fonctionnement émotionnel et relationnel inconscient, principalement à travers l'interprétation de ses manifestations actuelles dans son activité cognitive (cf. association libre) et dans la relation avec le psychothérapeute (transfert); et b) l'acquisition, à travers les expériences vécues dans la relation effective avec le thérapeute, de capacités d'autorégulation émotionnelles (ex. mentalisation) auparavant déficientes ou insuffisantes, qui vont lui permettre de modifier son fonctionnement émotionnel.

Dans la deuxième partie de notre mémoire, intitulée *explanans* (ce qui explique), nous avons présenté le cadre métathéorique du principe de l'énergie libre ainsi que les modèles théoriques intégratifs qui en découlent et qui ont été développés pour expliquer les dynamiques multiéchelles du fonctionnement neurocognitif humain. Ce cadre conceptuel et théorique nous a permis de nous former un modèle général nous permettant de nous représenter l'ensemble des échelles d'organisation des systèmes neurocognitifs humains qui sont pertinentes pour comprendre les phénomènes en jeu au sein de la psychothérapie psychodynamique.

Selon le principe de l'énergie libre (FEP) la manière dont les systèmes physiques particuliers que sont les systèmes vivants accomplissent l'exploit de rester en vie, c'est-à-dire, de maintenir leur organisation à travers le temps, consiste à travailler activement pour réduire la dispersion ou le désordre de leurs états sensoriels et physiologiques. Plus précisément, ils réalisent cela en minimisant leur *énergie libre variationnelle*, une quantité informationnelle qui encode la limite supérieure de l'entropie de leurs échanges sensoriels avec leur environnement (Friston, 2010). Le FEP affirme que pour qu'un système résiste à l'érosion entropique et se maintienne dans un ensemble d'états limité (pour posséder une homéostasie généralisée) il doit instancier un modèle statistique et causal de sa relation avec son éconiche (Ramstead *et al.*, 2017). Développé à partir du FEP, le modèle ontologique de la neuroéthologie variationnelle nous a permis de nous représenter les processus qui définissent les différentes échelles spatiotemporelles de l'auto-organisation des systèmes neurocognitifs humains comme différents moyens mécaniquement hétérogènes de minimiser l'énergie libre de leurs états internes (Ramstead *et al.*, 2017). Les outils conceptuels offerts par la neuroéthologie variationnelle nous ont permis de formuler une contrainte épistémique selon laquelle notre tentative d'explication de la psychothérapie psychodynamique devrait non seulement expliciter le ou les niveaux d'organisation des systèmes neurocognitifs humains impliqués dans cette pratique, mais aussi la manière dont les mécanismes ou les processus qui les définissent contribuent à la minimisation de leur énergie libre.

Ensuite, les approches prédictives nous ont fourni une description des mécanismes à travers lesquels le cerveau humain réalise le processus de minimisation de l'énergie libre : il incarne une machine à inférence qui cherche à prédire les entrées sensorielles provenant du milieu interne (intéroception) et externe (extéroception) à l'organisme auquel il appartient en générant des modèles probabilistes de leurs causes distales (Hohwy, 2013). Ici, la fonction du cerveau est définie comme la réalisation de l'allostase de l'organisme, qui correspond au processus dynamique de maintien des

états internes de ce dernier dans les normes homéostatiques définies par son phénotype (Barrett, 2016 ; Sterling, 2012). Nous avons décidé d'endosser la théorie des émotions de Barrett (2016), qui affirme que la manière principale dont le cerveau humain optimise le processus de réalisation l'allostasie de l'organisme auquel il appartient consiste à générer et à implémenter des *concepts émotionnels* : des modèles prédictifs multisensoriels qui expliquent les erreurs de prédiction intéroceptives, extéroceptives et proprioceptive afin générer une simulation dynamique permettant de prédire et de réguler l'état de l'organisme par rapport à sa situation dans son environnement, non seulement dans le présent, mais aussi dans le futur. Grâce au paradigme émergent de la neuropsychologie des grands réseaux, nous avons pu comprendre que la génération et l'implémentation des concepts émotionnels, de même que les autres formes de cognition de haut niveau qui sont impliquées dans la psychothérapie psychodynamique (ex. mentalisation), sont fort probablement sous-tendue par l'activité les échelles supérieures de la hiérarchie cérébrale, qui supervisent et coordonnent l'activité de l'ensemble du cerveau.

Pour finir, le « Skillfull intentionality framework » (Rietveld *et al.*, 2018) nous a permis de réaliser que l'échelle appropriée à l'étude des formes de cognition impliquées dans la psychothérapie psychodynamique est probablement celle du système cerveau-corps-environnement. Il nous a permis de comprendre le processus cérébral de minimisation de l'erreur de prédiction comme un mécanisme ayant pour but l'implémentation d'un couplage structurel robuste entre l'organisme et son environnement. Afin d'instaurer ce couplage structurel, l'organisme n'incarnerait pas un modèle de la structure causale de l'environnement en soi, mais plutôt un modèle de la structure dynamique d'un paysage *d'affordances*, des possibilités d'actions issues de l'interaction entre les habiletés d'un agent cognitif et les attributs pertinents de son environnement matériel, des actions ayant pour but de réaliser l'allostasie de l'organisme. Enfin, le modèle des affordances culturelles (Ramstead *et al.* 2016) nous ont offert une compréhension des mécanismes à travers lesquels les systèmes

neurocognitifs humains acquièrent la capacité de s'engager dans le paysage d'affordances leur permettant d'instaurer un couplage structurel avec l'échelle socioculturelle de leur niche écologique. Ce mécanisme correspondrait à l'engagement dans des régimes d'attention partagée, des pratiques collectives qui signalent la saillance (ou l'inverse) de certaines formes de stimuli et de situations.

Dans la troisième partie de notre mémoire intitulée *explication*, nous avons intégré des travaux récents réalisés dans le cadre des approches théoriques avancées dans notre *explanans* afin étayer la formulation d'hypothèses visant à identifier : a) le ou les niveaux d'auto-organisation spécifiques des systèmes cognitifs humains impliqués dans la pratique de la psychothérapie psychodynamique ; et b) les mécanismes à travers lesquels la psychothérapie psychodynamique participe à la modification de l'organisation de ce(s) niveau(x) permettant l'amélioration du fonctionnement cognitif et de la santé mentale des individus. Le modèle explicatif issu du travail réalisé dans cette troisième partie est résumé dans les paragraphes qui suivent.

La pratique de la psychothérapie psychodynamique repose sur l'engagement avec un niveau particulier de l'organisation des systèmes neurocognitifs humains (Carhart-Harris et Friston, 2010 ; Connolly et van Deventer, 2017 ; Ramstead *et al.*, 2017), dont l'existence et le fonctionnement sont déterminés par le couplage structurel intime qui lie un être humain avec son éconiche socioculturelle (Bruin *et al.*, 2018 ; Ramstead *et al.*, 2016), et qui s'instaure par le biais du processus collaboratif de réalisation de l'allostasie d'un individu à travers les relations d'attachement qui l'unissent avec ses donneurs de soins primaires durant l'ontogénie (Atzil *et al.*, 2018 ; Atzil et Barrett, 2017 ; Cittern *et al.*, 2018 ; Fotopoulou et Tsakiris, 2017). Plus précisément, les interactions récurrentes de communication coopérative visant à la modulation des états homéostatiques et affectifs d'un enfant humain mènent, à travers la synchronisation progressive des systèmes prédictifs incarnés par les cerveaux de ce dernier et des adultes qui s'en occupent, à l'émergence chez celui-ci d'une nouvelle échelle

d'organisation neurocognitive qui implémente la génération d'un modèle commun du phénotype et de la niche écologique qu'il partage avec ses congénères (Friston et Frith, 2015 ; Vasil *et al.*, 2020). Les processus cognitifs qui définissent ce niveau d'organisation correspondent à une suite d'habiletés de conceptualisation métacognitive (de mentalisation) offrant à l'individu la capacité de modéliser les processus causaux qui définissent l'échelle cognitive de l'organisation et de l'activité des organismes humains (Barrett, 2016 ; Fonagy et al., 2018 ; Moutoussis et al., 2014).

L'apprentissage de ces habiletés est sous-tendu par le développement des échelles supérieures de la hiérarchie prédictive incarnée par le cerveau humain, et en particulier du *réseau du mode par défaut*, dont l'activité génère et implémente des concepts métacognitifs, des modèles prédictifs d'un niveau d'abstraction élevée qui expliquent les dynamiques du fonctionnement affectif et comportemental (c.-à-d. émotionnel) de l'individu et de ses congénères en modélisant les processus cognitifs qui en sont responsables (Barrett et Satpute, 2013 ; Carhart-Harris et Friston, 2010 ; Luyten et Fonagy, 2015). Le processus de conceptualisation métacognitive réalisé par l'activité de ce réseau sous-tendrait les capacités d'autorégulation émotionnelle et de cognition sociale qui permettent à un être humain d'optimiser le processus de réalisation de son allostasie à et d'instaurer un couplage structurel robuste et intime avec son éconiche socioculturelle.

La psychothérapie psychodynamique correspond à une tentative de recréer les conditions naturelles de l'apprentissage des habiletés de conceptualisation métacognitives qui sous-tendent les capacités d'autorégulation émotionnelle et de cognition sociale d'un être humain, et ce afin de faciliter leur modification. Ces conditions correspondent à des relations interpersonnelles au sein desquelles des donneurs de soins cognitifs régulent et façonnent les capacités de conceptualisation métacognitive d'un individu, à travers des interactions récurrentes de communication coopérative qui mènent les dynamiques de l'organisation des échelles supérieures du

système neurocognitif de ce dernier à se synchroniser progressivement avec les leurs (Fonagy *et al.*, 2018 ; Vasil *et al.*, 2020). Afin de favoriser l'acquisition de capacités d'autorégulation émotionnelle et de cognition sociale fonctionnelles et adaptatives, la psychothérapie psychodynamique est pratiquée par un donneur de soins cognitifs aux caractéristiques spécifiques, un professionnel ayant suivi une formation afin de développer de très bonnes habiletés de conceptualisation métacognitive, et que l'on appelle un psychothérapeute.

En conclusion, la psychothérapie psychodynamique peut être définie comme une pratique culturelle partagée de communication coopérative au sein de laquelle un psychothérapeute et son client s'engagent dans une activité de co-conceptualisation métacognitive des processus cognitifs qui sous-tendent le fonctionnement émotionnel et interpersonnel de ce dernier (Gendron et Barrett, 2018 ; Vasil *et al.*, 2020). Cette pratique vise à l'amélioration des habiletés de conceptualisation métacognitive du client, ainsi qu'à entraîner ce dernier à l'exercice habile des capacités d'autorégulation émotionnelle et de cognition sociale qu'elles sous-tendent, tout ceci dans le but de favoriser la réalisation de son allostasie à travers l'instauration d'un couplage structurel robuste et intime avec son éconiche socioculturelle.

Nous pensons que le modèle théorique proposé ci-dessus constitue une tentative réussie de formuler une explication des phénomènes en jeu au sein de la psychothérapie psychodynamique qui soit compatible avec les intuitions fondamentales de la psychanalyse freudienne et de la psychologie psychodynamique, tout en étant réductible à des théories scientifiques actuelles, que ce soit en psychologie, en neurosciences cognitives, en anthropologie cognitive, etc. Autrement dit, nous pensons avoir réussi à proposer une défense adéquate de la thèse principale de notre mémoire de maîtrise. Cependant, le modèle avancé ici est essentiellement composé d'hypothèses issues de l'intégration de différents modèles théoriques, ce qui implique qu'il demande un processus d'investigation empirique afin de pouvoir être validé. En guise

d'ouverture, nous allons formuler ici quelques propositions sur la manière dont il pourrait être employé afin de guider la recherche empirique sur le fonctionnement et sur l'efficacité de la psychothérapie.

Premièrement, notre modèle fait l'hypothèse que l'efficacité thérapeutique de la psychothérapie psychodynamique est basée sur une amélioration des habiletés de conceptualisation métacognitive (mentalisation) des individus qui y participent. Pour tester cette hypothèse, il faudrait utiliser des outils de mesure clinique permettant d'évaluer les capacités de mentalisation des personnes suivant une psychothérapie psychodynamique de manière à vérifier que cette pratique mène bien à leur amélioration (Beaulieu-Pelletier *et al.*, 2013 ; Dimitrijević *et al.*, 2018 ; Fonagy *et al.*, 2016). Deuxièmement, notre modèle est basé sur l'idée selon laquelle les capacités de mentalisation d'un individu dépendent de l'activité du réseau du mode par défaut (Denny *et al.*, 2012 ; Frith *et al.*, 2003 ; Hyatt *et al.*, 2015 ; Lombardo *et al.*, 2009 ; Spreng *et al.*, 2008 ; Spunt *et al.*, 2010, 2015). Ainsi, une prédiction que nous pourrions émettre à partir de notre modèle serait que l'engagement prolongé dans la psychothérapie psychodynamique devrait mener à une modification significative de la structure et de la connectivité fonctionnelle du réseau du mode par défaut des individus qui y participent. De plus, le degré de cette modification devrait être corrélé avec l'amélioration des capacités de mentalisation, ainsi qu'avec une amélioration du bien-être psychologique des individus. Cette prédiction pourrait être testée en utilisant des techniques d'imagerie cérébrale afin de mesurer la structure du réseau du mode par défaut avant et après la psychothérapie. Enfin, afin de capturer de manière plus fine les dynamiques de l'interaction entre les systèmes neurocognitifs du client et du psychothérapeute qui ont lieu au sein de la psychothérapie psychodynamique, la technique de l'hyperscanning (Koike *et al.*, 2015), qui consiste à enregistrer simultanément l'activité cérébrale de plusieurs individus, pourrait être employée.

BIBLIOGRAPHIE

Agosta, L. (2010). Empathy as Vicarious Introspection in Psychoanalysis. Dans L. Agosta (dir.), *Empathy in the Context of Philosophy* (p. 132-157). London : Palgrave Macmillan UK. doi: 10.1057/9780230275249_7

Ainsworth, M. D. S., Blehar, M. C., Waters, E. et Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation* (p. xviii, 391). Oxford, England : Lawrence Erlbaum.

Amft, M., Bzdok, D., Laird, A. R., Fox, P. T., Schilbach, L. et Eickhoff, S. B. (2015). Definition and characterization of an extended social-affective default network. *Brain Structure and Function*, 220(2), 1031-1049. doi: 10.1007/s00429-013-0698-0

Amodio, D. M. et Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, 7(4), 268-277. doi: 10.1038/nrn1884

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D. et Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience*, 2(11), 1032-1037. doi: 10.1038/14833

Ardito, R. B. et Rabellino, D. (2011). Therapeutic Alliance and Outcome of Psychotherapy: Historical Excursus, Measurements, and Prospects for Research. *Frontiers in Psychology*, 2. doi: 10.3389/fpsyg.2011.00270

Arnold, S. E., Leuzinger-Bohleber, M. et Solms, M. (dir.). (2017). *The unconscious: a bridge between psychoanalysis and cognitive neuroscience*. London ; New York :

Routledge/Taylor & Francis Group.

Assoun, P.-L. (1990). *Introduction à l'épistémologie freudienne*. Paris : Payot.

Atzil, S. et Barrett, L. F. (2017). Social regulation of allostasis: Commentary on “Mentalizing homeostasis: The social origins of interoceptive inference” by Fotopoulou and Tsakiris. *Neuropsychoanalysis*, 19(1), 29-33. doi: 10.1080/15294145.2017.1295214

Atzil, S., Gao, W., Fradkin, I. et Barrett, L. F. (2018). Growing a social brain. *Nature Human Behaviour*, 2(9), 624-636. doi: 10.1038/s41562-018-0384-6

Badcock, P. B., Friston, K. J. et Ramstead, M. J. D. (2019a). The hierarchically mechanistic mind: A free-energy formulation of the human psyche. *Physics of Life Reviews*, 31, 104-121. doi: 10.1016/j.pprev.2018.10.002

Badcock, P. B., Friston, K., Ramstead, M. J. D., Ploeger, A. et Hohwy, J. (2019b). The hierarchically mechanistic mind: an evolutionary systems theory of the human brain, cognition, and behavior. *Cognitive, Affective, & Behavioral Neuroscience*, 19(6), 1319-1351. doi: 10.3758/s13415-019-00721-3

Baetens, K., Ma, N., Steen, J. et Van Overwalle, F. (2014). Involvement of the mentalizing network in social and non-social high construal. *Social Cognitive and Affective Neuroscience*, 9(6), 817-824. doi: 10.1093/scan/nst048

Baillargeon, R., Scott, R. M. et He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14(3), 110-118. doi: 10.1016/j.tics.2009.12.006

Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T., Probert Smith, P. J. et Woolrich, M. (2014). Fast transient networks in spontaneous human brain activity.

eLife, 3, e01867. doi: 10.7554/eLife.01867

Baldwin, S. A., Wampold et Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology*, 75(6), 842. doi: 10.1037/0022-006X.75.6.842

Barrett, L. F. (2016). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, nsw154. doi: 10.1093/scan/nsw154

Barrett, L. F. (2017). *How emotions are made: the secret life of the brain*. Boston : Houghton Mifflin Harcourt.

Barrett, L. F., Lewis, M. et Haviland-Jones, J. M. (2016). *Handbook of Emotions, Fourth Edition*. (s. l.) : Guilford Publications.

Barrett, L. F. et Satpute, A. B. (2013). Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology*, 23(3), 361-372. doi: 10.1016/j.conb.2012.12.012

Barrett, L. F. et Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, 16(7), 419-429. doi: 10.1038/nrn3950

Barsalou, L. W. (1999). Perceptual symbol systems. *BEHAVIORAL AND BRAIN SCIENCES*, 84.

Barsalou, L. W. (2009). *Simulation, situated conceptualization, and prediction*, 10. doi: 10.1098/rstb.2008.0319

Bartels, A. et Zeki, S. (2000). The neural basis of romantic love. *NeuroReport*, 11(17), 3829–3834. doi: 10.1097/00001756-200011270-00046

Beaulieu-Pelletier, G., Bouchard, M.-A. et Philippe, F. L. (2013). Mental States Task (MST): Development, Validation, and Correlates of a Self-Report Measure of Mentalization. *Journal of Clinical Psychology*, 69(7), 671-695. doi: 10.1002/jclp.21942

Beauregard, M., Lévesque, J. et Bourgouin, P. (2001). Neural Correlates of Conscious Self-Regulation of Emotion. *Journal of Neuroscience*, 21(18), RC165-RC165. doi: 10.1523/JNEUROSCI.21-18-j0001.2001

Beebe, B., Lachmann, F. et Jaffe, J. (1997). Mother–infant interaction structures and presymbolic self- and object representations. *Psychoanalytic Dialogues*, 7(2), 133-182. doi: 10.1080/10481889709539172

Beeghly, M. et Cicchetti, D. (1994). Child maltreatment, attachment, and the self system: Emergence of an internal state lexicon in toddlers at high social risk. *Development and Psychopathology*, 6(1), 5-30. doi: 10.1017/S095457940000585X

Berkovich-Ohana, A., Glicksohn, J. et Goldstein, A. (2012). Mindfulness-induced changes in gamma band activity – Implications for the default mode network, self-reference and attention. *Clinical Neurophysiology*, 123(4), 700-710. doi: 10.1016/j.clinph.2011.07.048

Bertalanffy, L. van. (2003). *General system theory: foundations, development, applications* (Rev. ed., 14. paperback print). New York : Braziller.

Bianchi-Demicheli, F. et Ortigue, S. (2007). Toward an understanding of the cerebral

substrates of woman's orgasm. *Neuropsychologia*, 45(12), 2645-2659. doi: 10.1016/j.neuropsychologia.2007.04.016

Binder, J. R., Desai, R. H., Graves, W. W. et Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex*, 19(12), 2767-2796. doi: 10.1093/cercor/bhp055

Bion, W. R. (1984). *Learning from Experience* (Reprint edition). London : Routledge.

Bluhm, R. L., Clark, C. R., McFarlane, A. C., Moores, K. A., Shaw, M. E. et Lanius, R. A. (2011). Default network connectivity during a working memory task. *Human Brain Mapping*, 32(7), 1029-1035. doi: 10.1002/hbm.21090

Boesch, C. (2020). Mothers, environment, and ontogeny affect cognition. *Animal Behavior and Cognition*, 7(3), 474-489. doi: 10.26451/abc.07.03.13.2020

Bowlby, J. (1999). *Attachment and loss* (2nd ed). New York : Basic Books.

Bracken, B. A. (1996). *Handbook of self-concept: Developmental, social, and clinical considerations* (p. xvi, 539). Oxford, England : John Wiley & Sons.

Bremner, J. D., Staib, L. H., Kaloupek, D., Southwick, S. M., Soufer, R. et Charney, D. S. (1999). Neural correlates of exposure to traumatic pictures and sound in Vietnam combat veterans with and without posttraumatic stress disorder: a positron emission tomography study. *Biological Psychiatry*, 45(7), 806-816. doi: 10.1016/S0006-3223(98)00297-2

Bressler, S. L. et Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences*, 14(6), 277-290. doi:

10.1016/j.tics.2010.04.004

Bressler, S. L. et Tognoli, E. (2006). Operational principles of neurocognitive networks. *International Journal of Psychophysiology*, 60(2), 139-148. doi: 10.1016/j.ijpsycho.2005.12.008

Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. *Metacognition, Motivation, and Understanding*. Récupéré de <https://ci.nii.ac.jp/naid/10020975585/>

Brown, H., Adams, R. A., Parees, I., Edwards, M. et Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing*, 14(4), 411-427. doi: 10.1007/s10339-013-0571-3

Brown, H. et Friston, K. (2013). The functional anatomy of attention: a DCM study. *Frontiers in Human Neuroscience*, 7. doi: 10.3389/fnhum.2013.00784

Brudzynski, S. M. (2013). The Ascending Mesolimbic Cholinergic System—A Specific Division of the Reticular Activating System Involved in the Initiation of Negative Emotional States. *Journal of Molecular Neuroscience*. doi: 10.1007/s12031-013-0179-1

Bruin, L. de, Newen, A. et Gallagher, S. (dir.). (2018). *The Oxford Handbook of 4E Cognition*. Oxford : Oxford University Press.

Bruineberg, J. et Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, 8. doi: 10.3389/fnhum.2014.00599

Bufill, E., Agustí, J. et Blesa, R. (2011). Human neoteny revisited: The case of synaptic

plasticity. *American Journal of Human Biology*, 23(6), 729-739. doi: 10.1002/ajhb.21225

Burston, D. (1986). The Cognitive and Dynamic Unconscious. *Contemporary Psychoanalysis*, 22(1), 133-157. doi: 10.1080/00107530.1986.10746119

Cacioppo, J. T., Berntson, G. G., Adolphs, R., Carter, C. S., McClintock, M. K., Meaney, M. J., ... Taylor, S. E. (2002). *Foundations in Social Neuroscience*. (s. l.) : MIT Press.

Campbell, L. F., Norcross, J. C., Vasquez, M. J. T. et Kaslow, N. J. (2013). Recognition of psychotherapy effectiveness: The APA resolution. *Psychotherapy*, 50(1), 98. doi: 10.1037/a0031817

Carhart-Harris, R. L. et Friston, K. (2010). The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas. *Brain*, 133(4), 1265-1283. doi: 10.1093/brain/awq010

Carpendale, J. et Lewis, C. (2006). *How children develop social understanding* (p. xii, 311). Malden : Blackwell Publishing.

Cassidy, J. et Shaver, P. R. (2002). *Handbook of Attachment: Theory, Research, and Clinical Applications*. (s. l.) : Rough Guides.

Castonguay, L. G., Castonguay, L. G. et Beutler, L. E. (2006). *Principles of Therapeutic Change that Work*. (s. l.) : Oxford University Press.

Chen, M. L. et Waxman, S. R. (2013). "Shall we blick?": Novel words highlight actors' underlying intentions for 14-month-old infants. *Developmental Psychology*, 49(3), 426-431. doi: 10.1037/a0029486

Cisek, P. et Kalaska, J. F. (2010). Neural Mechanisms for Interacting with a World Full of Action Choices. *Annual Review of Neuroscience*, 33(1), 269-298. doi: 10.1146/annurev.neuro.051508.135409

Cittern, D., Nolte, T., Friston, K. et Edalat, A. (2018). Intrinsic and extrinsic motivators of attachment under active inference. *PLOS ONE*, 35. doi: 10.1371/journal.pone.0193955

Clark, A. (2013). The many faces of precision (Replies to commentaries on “Whatever next? Neural prediction, situated agents, and the future of cognitive science”). *Frontiers in Psychology*, 4. doi: 10.3389/fpsyg.2013.00270

Cocchi, L., Zalesky, A., Fornito, A. et Mattingley, J. B. (2013). Dynamic cooperation and competition between brain systems during cognitive control. *Trends in Cognitive Sciences*, 17(10), 493-501. doi: 10.1016/j.tics.2013.08.006

Connolly, P. et van Deventer, V. (2017). Hierarchical Recursive Organization and the Free Energy Principle: From Biological Self-Organization to the Psychoanalytic Mind. *Frontiers in Psychology*, 8. doi: 10.3389/fpsyg.2017.01695

Constant, A., Clark, A., Kirchhoff, M. et Friston, K. (2019). Extended active inference: constructing predictive cognition beyond skulls. *Mind and Language*. Récupéré de <http://sro.sussex.ac.uk/id/eprint/88369/>

Coutinho, J. F., Fernandesl, S. V., Soares, J. M., Maia, L., Gonçalves, Ó. F. et Sampaio, A. (2016). Default mode network dissociation in depressive and anxiety states. *Brain Imaging and Behavior*, 10(1), 147-157. doi: 10.1007/s11682-015-9375-7

Craver, C. et Tabery, J. (2019). Mechanisms in Science. Dans E. N. Zalta (dir.), *The*

Stanford Encyclopedia of Philosophy (Summer 2019). Metaphysics Research Lab, Stanford University. Récupéré de <https://plato.stanford.edu/archives/sum2019/entries/science-mechanisms/>

Crits-Christoph, P., Gibbons, M. B. C., Hamilton, J., Ring-Kurtz, S. et Gallop, R. (2011). The dependability of alliance assessments: The alliance–outcome correlation is larger than you might think. *Journal of Consulting and Clinical Psychology*, 79(3), 267-278. doi: 10.1037/a0023668

Crone, E. A., Wendelken, C., Donohue, S. E. et Bunge, S. A. (2006). Neural Evidence for Dissociable Components of Task-switching. *Cerebral Cortex*, 16(4), 475-486. doi: 10.1093/cercor/bhi127

Damasio, A. R. (2018). *The Strange Order of Things: Life, Feeling, and the Making of Cultures*. (s. l.) : Vintage.

Dastjerdi, M., Foster, B. L., Nasrullah, S., Rauschecker, A. M., Dougherty, R. F., Townsend, J. D., ... Parvizi, J. (2011). Differential electrophysiological response during rest, self-referential, and non–self-referential tasks in human posteromedial cortex. *Proceedings of the National Academy of Sciences*, 108(7), 3023-3028. doi: 10.1073/pnas.1017098108

Davey, C. G., Pujol, J. et Harrison, B. J. (2016). Mapping the self in the brain’s default mode network. *NeuroImage*, 132, 390-397. doi: 10.1016/j.neuroimage.2016.02.022

Davidson, R. J., Putnam, K. M. et Larson, C. L. (2000). Dysfunction in the Neural Circuitry of Emotion Regulation--A Possible Prelude to Violence. *Science*, 289(5479), 591-594. doi: 10.1126/science.289.5479.591

De Jesus, P. (2016). Autopoietic enactivism, phenomenology and the deep continuity between life and mind. *Phenomenology and the Cognitive Sciences*, 15(2), 265-289. doi: 10.1007/s11097-015-9414-2

Dehaene, S. (2014). *Le code de la conscience*. Paris : O. Jacob.

Denny, B. T., Kober, H., Wager, T. D. et Ochsner, K. N. (2012). A Meta-analysis of Functional Neuroimaging Studies of Self- and Other Judgments Reveals a Spatial Gradient for Mentalizing in Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience*, 24(8), 1742-1752. doi: 10.1162/jocn_a_00233

Dimaggio, G., Lysaker, P. H., Carcione, A., Nicolò, G. et Semerari, A. (2008). Know yourself and you shall know the other... to a certain extent: Multiple paths of influence of self-reflection on mindreading. *Consciousness and Cognition*, 17(3), 778-789. doi: 10.1016/j.concog.2008.02.005

Dimitrijević, A., Hanak, N., Dimitrijević, A. A. et Marjanović, Z. J. (2018). The Mentalization Scale (MentS): A Self-Report Measure for the Assessment of Mentalizing Capacity. *Journal of Personality Assessment*, 100(3), 268-280. doi: 10.1080/00223891.2017.1310730

Dougherty, D. D., Rauch, S. L., Deckersbach, T., Marci, C., Loh, R., Shin, L. M., ... Fava, M. (2004). Ventromedial Prefrontal Cortex and Amygdala Dysfunction During an Anger Induction Positron Emission Tomography Study in Patients With Major Depressive Disorder With Anger Attacks. *Archives of General Psychiatry*, 61(8), 795. doi: 10.1001/archpsyc.61.8.795

Dragioti, E., Dimoliatis, I., Fountoulakis, K. N. et Evangelou, E. (2015). A systematic appraisal of allegiance effect in randomized controlled trials of psychotherapy. *Annals*

of General Psychiatry, 14. doi: 10.1186/s12991-015-0063-1

Duncan, S. et Barrett, L. F. (2007). Affect is a form of cognition: A neurobiological analysis. *Cognition and Emotion, 21(6)*, 1184-1211. doi: 10.1080/02699930701437931

Dunsworth, H. M., Warrener, A. G., Deacon, T., Ellison, P. T. et Pontzer, H. (2012). Metabolic hypothesis for human altriciality. *Proceedings of the National Academy of Sciences, 109(38)*, 15212-15216. doi: 10.1073/pnas.1205282109

Eagle, M. N. (2011). *From classical to contemporary psychoanalysis: a critique and integration.* New York, NY : Routledge.

Ebbesson, S. O. E. (1980). The parcellation theory and its relation to interspecific variability in brain organization, evolutionary and ontogenetic development, and neuronal plasticity. *Cell and Tissue Research, 213(2)*, 179-212. doi: 10.1007/BF00234781

Edelman, G. M. (1992). *Bright air, brilliant fire: On the matter of the mind* (p. xvi, 280). New York, NY, US : Basic Books.

Eriksen, C. W. et Kuehe, J. L. (1956). Avoidance conditioning of verbal behavior without awareness: A paradigm of repression. *The Journal of Abnormal and Social Psychology, 53(2)*, 203-209. doi: 10.1037/h0040008

Etkin, A., Egner, T. et Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Sciences, 15(2)*, 85-93. doi: 10.1016/j.tics.2010.11.004

Fairbairn, W. R. (1986). *Psychoanalytic studies of the personality* (8. impr., repr).

London : Routledge [and] Kegan Paul.

Falkenström, F. et Larsson, M. H. (2017). The Working Alliance: From Global Outcome Prediction to Micro-Analyses of Within-Session Fluctuations. *Psychoanalytic Inquiry*, 37(3), 167-178. doi: 10.1080/07351690.2017.1285186

Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I. et Posner, M. I. (2005). The activation of attentional networks. *NeuroImage*, 26(2), 471-479. doi: 10.1016/j.neuroimage.2005.02.004

Farmer, H. et Tsakiris, M. (2012). The Bodily Social Self: A Link Between Phenomenal and Narrative Selfhood. *Review of Philosophy and Psychology*, 3(1), 125-144. doi: 10.1007/s13164-012-0092-5

Feldman, H. et Friston, K. (2010). Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience*, 4. doi: 10.3389/fnhum.2010.00215

Feldman, R. (2015). The adaptive human parental brain: implications for children's social development. *Trends in Neurosciences*, 38(6), 387-399. doi: 10.1016/j.tins.2015.04.004

Ferry, A. L., Hespos, S. J. et Waxman, S. R. (2010). Categorization in 3- and 4-Month-Old Infants: An Advantage of Words Over Tones. *Child Development*, 81(2), 472-479. doi: 10.1111/j.1467-8624.2009.01408.x

Fivush, R. (2011). The Development of Autobiographical Memory. *Annual Review of Psychology*, 62(1), 559-582. doi: 10.1146/annurev.psych.121208.131702

Fivush, R. et Haden, C. A. (2003). *Autobiographical Memory and the Construction of A Narrative Self: Developmental and Cultural Perspectives*. (s. l.) : Psychology Press.

Flavell, J. H. (1976). Metacognitive Aspects of Problem Solving. *The nature of intelligence*. Récupéré de <https://ci.nii.ac.jp/naid/10021876052/>

Flückiger, C., Del Re, A. C., Wampold, Symonds, D. et Horvath, A. O. (2011). How central is the alliance in psychotherapy? A multilevel longitudinal meta-analysis. *Journal of Counseling Psychology*, 59(1), 10. doi: 10.1037/a0025749

Flückiger, C., Del Re, A. C., Wlodasch, D., Horvath, A. O., Solomonov, N. et Wampold, B. E. (2020). Assessing the alliance–outcome association adjusted for patient characteristics and treatment processes: A meta-analytic summary of direct comparisons. *Journal of Counseling Psychology*, No Pagination Specified-No Pagination Specified. doi: 10.1037/cou0000424

Fodor, J. (1986). *La modularité de l'esprit. Essai sur la psychologie des facultés*. Paris : Editions de Minuit.

Fonagy, P. (2015). The effectiveness of psychodynamic psychotherapies: An update. *World Psychiatry*, 14(2), 137-150. doi: 10.1002/wps.20235

Fonagy, P. et Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy*, 51(3), 372-380. doi: 10.1037/a0036505

Fonagy, P., Gergely, G., Jurist, E. L. et Target, M. (2018). *Affect regulation, mentalization, and the development of the self*. London New York : Routledge, Taylor & Francis Group.

Fonagy, P. et Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and Psychopathology*, 21(4), 1355-1381. doi: 10.1017/S0954579409990198

Fonagy, P. et Luyten, P. (2012). The multidimensional construct of mentalization and its relevance to understanding borderline personality disorder. Dans A. Fotopoulou, M. A. Conway et D. W. Pfaff (dir.), *From the couch to the lab: trends in psychodynamic neuroscience*. Oxford : Oxford University Press.

Fonagy, P., Luyten, P. et Allison, E. (2015). Epistemic Petrification and the Restoration of Epistemic Trust: A New Conceptualization of Borderline Personality Disorder and Its Psychosocial Treatment. *Journal of Personality Disorders*, 29(5), 575-609. doi: 10.1521/pedi.2015.29.5.575

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y.-W., Warren, F., Howard, S., ... Lowyck, B. (2016). Development and Validation of a Self-Report Measure of Mentalizing: The Reflective Functioning Questionnaire. *PLOS ONE*, 11(7), e0158678. doi: 10.1371/journal.pone.0158678

Fotopoulou, A., Pfaff, D. W. et Conway, M. A. (dir.). (2012). *From the couch to the lab: trends in psychodynamic neuroscience*. Oxford : Oxford University Press.

Fotopoulou, A. et Tsakiris, M. (2017). Mentalizing homeostasis: The social origins of interoceptive inference. *Neuropsychoanalysis*, 19(1), 3-28. doi: 10.1080/15294145.2017.1294031

Fox, M. D. et Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, 8(9), 700-711. doi: 10.1038/nrn2201

Fraley, R. C. et Roisman, G. I. (2019). The development of adult attachment styles: four lessons. *Current Opinion in Psychology*, 25, 26-30. doi: 10.1016/j.copsyc.2018.02.008

Frank, M. J. (2005). Dynamic Dopamine Modulation in the Basal Ganglia: A Neurocomputational Account of Cognitive Deficits in Medicated and Nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience*, 17(1), 51-72. doi: 10.1162/0898929052880093

Freud, A. (1992). *The Ego and the Mechanisms of Defence* (1 edition). (s. l.) : Routledge.

Freud, S. (1886). *Pre-psycho-analytic publications and unpublished drafts: 1886-1899*. London : Vintage.

Freud, S. (1954). Project for a scientific psychology. Dans M. Bonaparte, A. Freud, E. Kris, E. Mosbacher et J. Strachey (dir.), *The origins of psycho-analysis: Letters to Wilhelm Fliess, drafts and notes: 1887-1902* (p. 347-445). New York, NY, US : Basic Books. doi: 10.1037/11538-013

Freud, S. (1995). *New introductory lectures on psycho-analysis*. New York; London : W. W. Norton.

Freud, S. (2001). *Beyond the pleasure principle: group psychology and other works ; (1920-1922)*. London : Vintage.

Freud, S. (2010). *The interpretation of dreams*. New York : Basic Books A Member of the Perseus Books Group.

Freud, S., Breuer, J. et Berman, A. (2016). *Études sur l'hystérie*. (s. l. : n. é.).

Friedman, H. S. et Silver, R. C. (dir.). (2007). *Foundations of health psychology*. Oxford ; New York : Oxford University Press.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, 11(2), 127-138. doi: 10.1038/nrn2787

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P. et Pezzulo, G. (2017). Active Inference: A Process Theory. *Neural Computation*, 29(1), 1-49. doi: 10.1162/NECO_a_00912

Friston, K. et Frith, C. (2015). A Duet for one. *Consciousness and Cognition*, 36, 390-405. doi: 10.1016/j.concog.2014.12.003

Friston, K. J. (2000). The labile brain. I. Neuronal transients and nonlinear coupling. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355(1394), 215-236. doi: 10.1098/rstb.2000.0560

Friston, K., Kilner, J. et Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, 100(1), 70-87. doi: 10.1016/j.jphysparis.2006.10.001

Friston, K. et Klaas E., S. (2007). Free-energy and the brain. *Synthese*, 159(3), 417-458. doi: 10.1007/s11229-007-9237-y

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T. et Dolan, R. J. (2014). The anatomy of choice: dopamine and decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655), 20130481. doi: 10.1098/rstb.2013.0481

Friston, K., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., ... Bestmann, S. (2012). Dopamine, Affordance and Active Inference. *PLOS Computational Biology*, 8(1), e1002327. doi: 10.1371/journal.pcbi.1002327

Frith, C. D., Blakemore, S.-J. et Wolpert, D. M. (2000). Abnormalities in the awareness

and control of action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355(1404), 1771-1788. doi: 10.1098/rstb.2000.0734

Frith, C. d., Wolpert, D. m., Frith, U. et Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431), 459-473. doi: 10.1098/rstb.2002.1218

Fulgencio, L. (2005). Freud's metapsychological speculations. *The International Journal of Psychoanalysis*, 86(1), 99-123. doi: 10.1516/TNY1-6YT7-V37T-16LD

Gabbard, G. O. et Crocq, M.-A. (2010). *Psychothérapie psychodynamique: les concepts fondamentaux*. Issy-les-Moulineaux (Hauts-de-Seine) : Elsevier Masson. Récupéré de <http://site.ebrary.com/id/10537271>

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*, 4(1), 14-21. doi: 10.1016/S1364-6613(99)01417-5

Gao, W., Gilmore, J. H., Giovanello, K. S., Smith, J. K., Shen, D., Zhu, H. et Lin, W. (2011). Temporal and Spatial Evolution of Brain Network Topology during the First Two Years of Life. *PLoS ONE*, 6(9). doi: 10.1371/journal.pone.0025278

Gao, W., Gilmore, J. H., Shen, D., Smith, J. K., Zhu, H. et Lin, W. (2013). The Synchronization within and Interaction between the Default and Dorsal Attention Networks in Early Infancy. *Cerebral Cortex*, 23(3), 594-603. doi: 10.1093/cercor/bhs043

Gendron, M. et Barrett, L. F. (2018). Emotion Perception as Conceptual Synchrony.

Emotion Review, 10(2), 101-110. doi: 10.1177/1754073917705717

George, C. et Solomon, J. (2008). The caregiving system: A behavioral systems approach to parenting. Dans *Handbook of attachment: Theory, research, and clinical applications*, 2nd ed (p. 833-856). New York, NY, US : The Guilford Press.

Gergely, G. (2007). The social construction of the subjective self: The role of affect-mirroring, markedness, and ostensive communication in self-development. Dans *Developmental science and psychoanalysis: Integration and innovation* (p. 45-88). London, England : Karnac Books.

Gerlach, K. D., Spreng, R. N., Gilmore, A. W. et Schacter, D. L. (2011). Solving future problems: Default network and executive activity associated with goal-directed mental simulations. *NeuroImage*, 55(4), 1816-1824. doi: 10.1016/j.neuroimage.2011.01.030

Gibson, J. J. (2011). *The ecological approach to visual perception* (17th pr). New York : Psychology Press.

Glennan, S. (2017). *The new mechanical philosophy* (First edition). Oxford : Oxford University Press.

Goulden, N., Khusnulina, A., Davis, N. J., Bracewell, R. M., Bokde, A. L., McNulty, J. P. et Mullins, P. G. (2014). The salience network is responsible for switching between the default mode network and the central executive network: Replication from DCM. *NeuroImage*, 99, 180-190. doi: 10.1016/j.neuroimage.2014.05.052

Grafman, J., Schwab, K., Warden, D., Pridgen, A., Brown, H. R. et Salazar, A. M. (1996). Frontal lobe injuries, violence, and aggression: A report of the Vietnam Head Injury Study. *Neurology*, 46(5), 1231-1231. doi: 10.1212/WNL.46.5.1231

Granqvist, P. (2020). Attachment, culture, and gene-culture co-evolution: expanding the evolutionary toolbox of attachment theory. *Attachment & Human Development*, 0(0), 1-24. doi: 10.1080/14616734.2019.1709086

Greenson, R. R. (1978). *The technique and practice of psychoanalysis*. London : Hogarth P.

Greicius, M. D., Krasnow, B., Reiss, A. L. et Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100(1), 253-258. doi: 10.1073/pnas.0135058100

Grillner, S., Hellgren, J., Ménard, A., Saitoh, K. et Wikström, M. A. (2005). Mechanisms for selection of basic motor programs – roles for the striatum and pallidum. *Trends in Neurosciences*, 28(7), 364-370. doi: 10.1016/j.tins.2005.05.004

Gusnard, D. A., Akbudak, E., Shulman, G. L. et Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(7), 4259-4264. doi: 10.1073/pnas.071043098

Gweon, H., Dodell-Feder, D., Bedny, M. et Saxe, R. (2012). Theory of Mind Performance in Children Correlates With Functional Specialization of a Brain Region for Thinking About Thoughts. *Child Development*, 83(6), 1853-1868. doi: 10.1111/j.1467-8624.2012.01829.x

Hariri, A. R., Bookheimer, S. Y. et Mazziotta, J. C. (2000). Modulating emotional responses: effects of a neocortical network on the limbic system. *NeuroReport*, 11(1), 43–48. doi: 10.1097/00001756-200001170-00009

Hartmann, H. (1964). *Essays on Ego Psychology: Selected Problems in Psychoanalytic Theory*. (s. l.) : Hogarth Press.

Hasson, U. et Frith, C. D. (2016). Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1693), 20150366. doi: 10.1098/rstb.2015.0366

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S. et Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, 16(2), 114-121. doi: 10.1016/j.tics.2011.12.007

Hattie, J. (2014). *Self-Concept*. (s. l.) : Psychology Press.

Hayes, D. J. et Northoff, G. (2011). Identifying a Network of Brain Regions Involved in Aversion-Related Processing: A Cross-Species Translational Investigation. *Frontiers in Integrative Neuroscience*, 5. doi: 10.3389/fnint.2011.00049

Henrich, J. P. (2016). *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton : Princeton University Press.

Hervé, P.-Y., Razafimandimby, A., Jobard, G. et Tzourio-Mazoyer, N. (2013). A Shared Neural Substrate for Mentalizing and the Affective Component of Sentence Comprehension. *PLoS ONE*, 8(1). doi: 10.1371/journal.pone.0054400

Hesse, E. (2008). The Adult Attachment Interview: Protocol, method of analysis, and empirical studies. Dans *Handbook of attachment: Theory, research, and clinical applications*, 2nd ed (p. 552-598). New York, NY, US : The Guilford Press.

Heyes, C. M. (2018). *Cognitive gadgets: the cultural evolution of thinking*. Cambridge, Massachusetts London, England : The Belknap Press of Harvard University Press.

Heylighen, F. et Joslyn, C. A. (1992). *What is Systems Theory?* Récupéré de <http://pespmc1.vub.ac.be/SYSTHEOR.html>

Hinde, R. A. (1974). *Biological bases of human social behaviour* (p. xv, 462). New York, NY, US : McGraw-Hill.

Hobson, J. A., Pace-Schott, E. F. et Stickgold, R. (2000). Dreaming and the brain: Toward a cognitive neuroscience of conscious states. *Behavioral and Brain Sciences*, 23(6), 793-842. doi: 10.1017/S0140525X00003976

Hoffmann, M. (2020). Default Mode Network (Mentalizing) and Limbic Network for Explicit Episodic Memory Syndromes. Dans M. Hoffmann (dir.), *Clinical Mentation Evaluation: A Connectomal Approach to Rapid and Comprehensive Assessment* (p. 79-88). Cham : Springer International Publishing. doi: 10.1007/978-3-030-46324-3_8

Hohwy, J. (2013). *The predictive mind* (First edition). Oxford, United Kingdom ; New York, NY, United States of America : Oxford University Press.

Hohwy, J. (2016). The Self-Evidencing Brain. *Noûs*, 50(2), 259-285. doi: 10.1111/nous.12062

Hopkins, J. (2012). Psychoanalysis, representation, and neuroscience: The Freudian unconscious and the Bayesian brain. *From the couch to the lab: Trends in psychodynamic neuroscience*, 230–265. doi: 10.1093/med/9780199600526.003.0014

Hopper, J. W., Frewen, P. A., Kolk, B. A. van der et Lanius, R. A. (2007). Neural correlates of reexperiencing, avoidance, and dissociation in PTSD: Symptom

dimensions and emotion dysregulation in responses to script-driven trauma imagery. *Journal of Traumatic Stress*, 20(5), 713-725. doi: 10.1002/jts.20284

Hoskinson, K. R., Bigler, E. D., Abildskov, T. J., Dennis, M., Taylor, H. G., Rubin, K., ... Yeates, K. O. (2019). The mentalizing network and theory of mind mediate adjustment after childhood traumatic brain injury. *Social Cognitive and Affective Neuroscience*, 14(12), 1285-1295. doi: 10.1093/scan/nsaa006

Hrdy, S. B. (1999). *The Woman That Never Evolved*. Cambridge, Mass : Harvard University Press.

Hrdy, S. B. (2000). *Mother Nature: Maternal Instincts and How They Shape the Human Species*. New York : Ballantine Books.

Hrdy, S. B. (2009). *Mothers and others: the evolutionary origins of mutual understanding*. Cambridge, Mass : Belknap Press of Harvard University Press.

Hrdy, S. B. (2017, 28 juillet). Comes the Child before Man : How Cooperative Breeding and Prolonged Postweaning Dependence Shaped Human Potential. Dans *Hunter-Gatherer Childhoods*. doi: 10.4324/9780203789445-4

Hutto, D. D. et Satne, G. (2015). The Natural Origins of Content. *Philosophia*, 43(3), 521-536. doi: 10.1007/s11406-015-9644-0

Hyatt, C. J., Calhoun, V. D., Pearlson, G. D. et Assaf, M. (2015). Specific default mode subnetworks support mentalizing as revealed through opposing network recruitment by social and semantic fMRI tasks. *Human Brain Mapping*, 36(8), 3047-3063. doi: 10.1002/hbm.22827

Immordino-Yang, M. H. (2016). Emotion, Sociality, and the Brain's Default Mode

Network: Insights for Educational Practice and Policy. *Policy Insights from the Behavioral and Brain Sciences*, 3(2), 211-219. doi: 10.1177/2372732216656869

Joffily, M. et Coricelli, G. (2013). Emotional Valence and the Free-Energy Principle. *PLoS Computational Biology*, 9(6), e1003094. doi: 10.1371/journal.pcbi.1003094

Johnson, M. H. (1999). Ontogenetic constraints on neural and behavioral plasticity: Evidence from imprinting and face processing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 53(1), 77-91. doi: 10.1037/h0087301

Kahneman, D. (2013). *Thinking, Fast and Slow*. (s. l.) : Anchor Canada.

Kanai, R., Komura, Y., Shipp, S. et Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668), 20140169. doi: 10.1098/rstb.2014.0169

Kaplan-Solms, K. et Solms, M. (2002). *Clinical studies in neuro-psychoanalysis introduction to a depth neuropsychology*. New York : Karnac. Récupéré de <http://site.ebrary.com/id/10463983>

Karmali, F., Whitman, G. T. et Lewis, R. F. (2017). Bayesian optimal adaptation explains age-related human sensorimotor changes. *Journal of Neurophysiology*, 119(2), 509-520. doi: 10.1152/jn.00710.2017

Kihlstrom, J. F. (1987). The cognitive unconscious. *Science*, 237(4821), 1445-1452. doi: 10.1126/science.3629249

Kihlstrom, J. F. (1996). Perception without awareness of what is perceived, learning without awareness of what is learned. Dans *The science of consciousness:*

Psychological, neuropsychological and clinical reviews (p. 23-46). Florence, KY, US : Taylor & Frances/Routledge. doi: 10.4324/9780203360019_chapter_2

Kihlstrom, J. F. (2015). Dynamic versus Cognitive Unconscious. Dans *The Encyclopedia of Clinical Psychology* (p. 1-8). American Cancer Society. doi: 10.1002/9781118625392.wbecp275

Kirchhoff, M. D. (2018). Autopoiesis, free energy, and the life–mind continuity thesis. *Synthese*, 195(6), 2519-2540. doi: 10.1007/s11229-016-1100-6

Kirchhoff, M., Parr, T., Palacios, E., Friston, K. et Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *Journal of The Royal Society Interface*, 15(138), 20170792. doi: 10.1098/rsif.2017.0792

Klein, M. (1999). *Love, Guilt and Reparation: And Other Works 1921-1945*. New York : Free Press.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K. et Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage*, 42(2), 998-1031. doi: 10.1016/j.neuroimage.2008.03.059

Koechlin, E. et Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, 11(6), 229-235. doi: 10.1016/j.tics.2007.04.005

Kohut, H. (2009). *The restoration of the self*. Chicago : The University of Chicago Press.

Koike, T., Tanabe, H. C. et Sadato, N. (2015). Hyperscanning neuroimaging technique

to reveal the “two-in-one” system in social interactions. *Neuroscience Research*, 90, 25-32. doi: 10.1016/j.neures.2014.11.006

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G. et de Lange, F. P. (2016). Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Current Biology*, 26(3), 371-376. doi: 10.1016/j.cub.2015.12.038

Kovács, Á. M., Téglás, E. et Endress, A. D. (2010). The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. *Science*, 330(6012), 1830-1834. doi: 10.1126/science.1190792

Kovalzon, V. M. (2016). Ascending Reticular Activating System of the Brain. *Translational Neuroscience and Clinics*, 2(4), 275-285. doi: 10.18679/CN11-6030_R.2016.034

Krahé, C., Springer, A., Weinman, J. A. et Fotopoulou, A. (Katerina). (2013). The Social Modulation of Pain: Others as Predictive Signals of Salience – a Systematic Review. *Frontiers in Human Neuroscience*, 7. doi: 10.3389/fnhum.2013.00386

LaFreniere, P. J. (2000). *Emotional development: A biosocial perspective* (p. xx, 331). Belmont, CA, US : Wadsworth/Thomson Learning.

Lamm, C., Decety, J. et Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage*, 54(3), 2492-2502. doi: 10.1016/j.neuroimage.2010.10.014

Lanius, R. A., Williamson, P. C., Boksman, K., Densmore, M., Gupta, M., Neufeld, R. W. J., ... Menon, R. S. (2002). Brain activation during script-driven imagery induced dissociative responses in PTSD: a functional magnetic resonance imaging investigation.

Biological Psychiatry, 52(4), 305-311. doi: 10.1016/S0006-3223(02)01367-7

Laplanche, J. et Pontalis, J.-B. (2009). *Vocabulaire de la psychanalyse* (5. ed). Paris : Pr. Univ. de France.

Leichsenring, F., Hiller, W., Weissberg, M. et Leibing, E. (2006). Cognitive-Behavioral Therapy and Psychodynamic Psychotherapy: Techniques, Efficacy, and Indications. *American Journal of Psychotherapy*, 60(3), 233-259. doi: 10.1176/appi.psychotherapy.2006.60.3.233

Li, W., Mai, X. et Liu, C. (2014). The default mode network and social understanding of others: what do brain connectivity studies tell us. *Frontiers in Human Neuroscience*, 8. doi: 10.3389/fnhum.2014.00074

Lieberman, M. D. (2007). Social Cognitive Neuroscience: A Review of Core Processes. *Annual Review of Psychology*, 58(1), 259-289. doi: 10.1146/annurev.psych.58.110405.085654

Lindquist, K. A. et Barrett, L. F. (2012). A functional architecture of the human brain: emerging insights from the science of emotion. *Trends in Cognitive Sciences*, 16(11), 533-540. doi: 10.1016/j.tics.2012.09.005

Lindquist, K. A., MacCormack, J. K. et Shablack, H. (2015). The role of language in emotion: predictions from psychological constructionism. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.00444

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E. et Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, 35(3), 121-143. doi: 10.1017/S0140525X11000446

Livesley, D. W. J. (2007). An integrated approach to the treatment of personality disorder. *Journal of Mental Health, 16*(1), 131-148. doi: 10.1080/09638230601182086

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J. et Baron-Cohen, S. (2009). Shared Neural Circuits for Mentalizing about the Self and Others. *Journal of Cognitive Neuroscience, 22*(7), 1623-1635. doi: 10.1162/jocn.2009.21287

Lovero, K. L., Simmons, A. N., Aron, J. L. et Paulus, M. P. (2009). Anterior insular cortex anticipates impending stimulus significance. *NeuroImage, 45*(3), 976-983. doi: 10.1016/j.neuroimage.2008.12.070

Lupyan, G. et Clark, A. (2015). Words and the World: Predictive Coding and the Language-Perception-Cognition Interface. *Current Directions in Psychological Science, 24*(4), 279-284. doi: 10.1177/0963721415570732

Luyten, P. et Fonagy, P. (2015). The Neurobiology of Mentalizing. *Personality Disorders: Theory, Research, and Treatment, 6*(4), 366. doi: 10.1037/per0000117

Luyten, P. et Fonagy, P. (2018). The stress–reward–mentalizing model of depression: An integrative developmental cascade approach to child and adolescent depressive disorder based on the Research Domain Criteria (RDoC) approach. *Clinical Psychology Review, 64*, 87-98. doi: 10.1016/j.cpr.2017.09.008

Lyons-Ruth, K., Bronfman, E. et Parsons, E. (1999). Chapter IV. Maternal Frightened, Frightening, or Atypical Behavior and Disorganized Infant Attachment Patterns. *Monographs of the Society for Research in Child Development, 64*(3), 67-96. doi: 10.1111/1540-5834.00034

Lyons-Ruth, K., Bruschweiler-Stern, N., Harrison, A. M., Morgan, A. C., Nahum, J. P., Sander, L., ... Tronick, E. Z. (1998). Implicit relational knowing: Its role in development and psychoanalytic treatment. *Infant Mental Health Journal, 19*(3), 282-289. doi: 10.1002/(SICI)1097-0355(199823)19:3<282::AID-IMHJ3>3.0.CO;2-O

Macmillan, M. (2000). Nineteenth-century inhibitory theories of thinking: Bain, Ferrier, Freud (and Phineas Gage). *History of Psychology, 3*(3), 187. doi: 10.1037/1093-4510.3.3.187

Madigan, S., Moran, G. et Pederson, D. R. (2006). Unresolved states of mind, disorganized attachment relationships, and disrupted interactions of adolescent mothers and their infants. *Developmental Psychology, 42*(2), 293. doi: 10.1037/0012-1649.42.2.293

Mahler, M. S., Pine, F. et Bergman, A. (1975). *The Psychological Birth of the Human Infant: Symbiosis and Individuation*. (s. l.) : Maresfield Library.

Main, M., Kaplan, N. et Cassidy, J. (1985). Security in Infancy, Childhood, and Adulthood: A Move to the Level of Representation. *Monographs of the Society for Research in Child Development, 50*(1/2), 66-104. doi: 10.2307/3333827

Mars, R. B., Neubert, F.-X., Noonan, M. P., Sallet, J., Toni, I. et Rushworth, M. F. S. (2012). On the relationship between the “default mode network” and the “social brain”. *Frontiers in Human Neuroscience, 6*. doi: 10.3389/fnhum.2012.00189

Martins, B. et Mather, M. (2016). Default mode network and later-life emotion regulation: Linking functional connectivity patterns and emotional outcomes. Dans *Emotion, aging, and health* (p. 9-29). Washington, DC, US : American Psychological Association. doi: 10.1037/14857-002

Maturana, H. R. et Varela, F. J. (1980). *Autopoiesis and cognition: the realization of the living*. Dordrecht, Holland ; Boston : D. Reidel Pub. Co.

McAdams, D. P. (2013). The Psychological Self as Actor, Agent, and Author. *Perspectives on Psychological Science*, 8(3), 272-295. doi: 10.1177/1745691612464657

McKinnon, S. et Silverman, S. (2005). *Complexities: Beyond Nature and Nurture*. (s. l.) : University of Chicago Press.

Menon, V. (2011). Large-scale brain networks and psychopathology: a unifying triple network model. *Trends in Cognitive Sciences*, 15(10), 483-506. doi: 10.1016/j.tics.2011.08.003

Menon, V. et Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, 214(5), 655-667. doi: 10.1007/s00429-010-0262-0

Merleau-Ponty, M. (2009). *Phénoménologie de la perception*. Paris : Gallimard.

Messer, S. B. et Wampold. (2002). Let's Face Facts: Common Factors Are More Potent Than Specific Therapy Ingredients. *Clinical Psychology: Science and Practice*, 9(1), 21-25. doi: 10.1093/clipsy.9.1.21

Metcalfe, B. in the D. of P. J., Metcalfe, J. et Shimamura, A. P. (1994). *Metacognition: Knowing about Knowing*. (s. l.) : MIT Press.

Metzinger, T. (2004). *Being no one: the self-model theory of subjectivity* (1. paperback ed). Cambridge, Mass. : MIT Press.

- Mikulincer, M. et Shaver, P. R. (2012). An attachment perspective on psychopathology. *World Psychiatry, 11*(1), 11-15. doi: 10.1016/j.wpsyc.2012.01.003
- Milad, M. R. et Quirk, G. J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature, 420*(6911), 70-74. doi: 10.1038/nature01138
- Milad, M. R., Rauch, S. L., Pitman, R. K. et Quirk, G. J. (2006). Fear extinction in rats: Implications for human brain imaging and anxiety disorders. *Biological Psychology, 73*(1), 61-71. doi: 10.1016/j.biopsycho.2006.01.008
- Miller, E. K. et Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience, 24*(1), 167-202. doi: 10.1146/annurev.neuro.24.1.167
- Milnor, J. W. (2006). Attractor. *Scholarpedia, 1*(11), 1815. doi: 10.4249/scholarpedia.1815
- Mitchell, S. A. (1988). *Relational concepts in psychoanalysis: An integration* (p. ix, 326). Cambridge, MA, US : Harvard University Press.
- Mitchell, S. A. (2000). *Relationality: from attachment to intersubjectivity*. Hillsdale, NJ : Analytic Press.
- Molnar-Szakacs, I. et Uddin, L. Q. (2013). Self-Processing and the Default Mode Network: Interactions with the Mirror Neuron System. *Frontiers in Human Neuroscience, 7*. doi: 10.3389/fnhum.2013.00571
- Moore, C. et Dunham, P. J. (dir.). (1995). *Joint attention: its origins and role in development*. Hillsdale, N.J : Lawrence Erlbaum Associates.

Moran, J. M., Kelley, W. M. et Heatherton, T. F. (2013). What Can the Organization of the Brain's Default Mode Network Tell us About Self-Knowledge? *Frontiers in Human Neuroscience*, 7. doi: 10.3389/fnhum.2013.00391

Moustafa, A. A., Sherman, S. J. et Frank, M. J. (2008). A dopaminergic basis for working memory, learning and attentional shifting in Parkinsonism. *Neuropsychologia*, 46(13), 3144-3156. doi: 10.1016/j.neuropsychologia.2008.07.011

Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J. et Friston, K. (2014a). Bayesian inferences about the self (and others): A review. *Consciousness and Cognition*, 25, 67-76. doi: 10.1016/j.concog.2014.01.009

Moutoussis, M., Trujillo-Barreto, N. J. P., El-Deredy, W., Dolan, R. et Friston, K. (2014b). A formal model of interpersonal inference. *Frontiers in Human Neuroscience*, 8. doi: 10.3389/fnhum.2014.00160

Müller, N. G. et Knight, R. T. (2006). The functional neuroanatomy of working memory: Contributions of human brain lesion studies. *Neuroscience*, 139(1), 51-58. doi: 10.1016/j.neuroscience.2005.09.018

Munder, T., Flückiger, C., Leichsenring, F., Abbass, A. A., Hilsenroth, M. J., Luyten, P., ... Wampold. (2018). Is psychotherapy effective? A re-analysis of treatments for depression. *Epidemiology and Psychiatric Sciences*, 1-7. doi: 10.1017/S2045796018000355

Muscattell, K. A., Morelli, S. A., Falk, E. B., Way, B. M., Pfeifer, J. H., Galinsky, A. D., ... Eisenberger, N. I. (2012). Social status modulates neural activity in the mentalizing network. *NeuroImage*, 60(3), 1771-1777. doi: 10.1016/j.neuroimage.2012.01.080

Nelson, S. M., Dosenbach, N. U. F., Cohen, A. L., Wheeler, M. E., Schlaggar, B. L. et Petersen, S. E. (2010). Role of the anterior insula in task-level control and focal attention. *Brain Structure and Function*, 214(5-6), 669-680. doi: 10.1007/s00429-010-0260-2

Norcross, J. C. (dir.). (2019). *Psychotherapy relationships that work* (Third edition). New York, NY : Oxford University Press.

Ochsner, K. N. et Gross, J. J. (2008). Cognitive Emotion Regulation: Insights From Social Cognitive and Affective Neuroscience. *Current Directions in Psychological Science*, 17(2), 153-158. doi: 10.1111/j.1467-8721.2008.00566.x

Páez-Rondón, O., Aldana, E., Dickens, J. et Otálora-Luna, F. (2018). Ethological description of a fixed action pattern in a kissing bug (Triatominae): vision, gustation, proboscis extension and drinking of water and guava. *Journal of Ethology*, 36(2), 107-116. doi: 10.1007/s10164-018-0547-y

Palacios, E. R., Razi, A., Parr, T., Kirchhoff, M. et Friston, K. (2020). On Markov blankets and hierarchical self-organisation. *Journal of Theoretical Biology*, 486, 110089. doi: 10.1016/j.jtbi.2019.110089

Palmer, C. J., Seth, A. K. et Hohwy, J. (2015). The felt presence of other minds: Predictive processing, counterfactual predictions, and mentalising in autism. *Consciousness and Cognition*, 36, 376-389. doi: 10.1016/j.concog.2015.04.007

Panksepp, J. (2004). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. (s. l.) : Oxford University Press.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and

humans. *Consciousness and Cognition*, 14(1), 30-80. doi: 10.1016/j.concog.2004.10.004

Panksepp, J. (2006). Emotional endophenotypes in evolutionary psychiatry. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 30(5), 774-784. doi: 10.1016/j.pnpbp.2006.01.004

Panksepp, J. (2010). Affective neuroscience of the emotional BrainMind: evolutionary perspectives and implications for understanding depression. *Dialogues in Clinical Neuroscience*, 12(4), 533-545.

Parvizi, J. et Damasio, A. (2001). Consciousness and the brainstem. *Cognition*, 79(1), 135-160. doi: 10.1016/S0010-0277(00)00127-X

Penfield, W. et Jasper, H. (1954). *Epilepsy and the functional anatomy of the human brain* (p. xv, 896). Oxford, England : Little, Brown & Co.

Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 781-795. doi: 10.1098/rstb.2005.1631

Pezzulo, G., Rigoli, F. et Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, 134, 17-35. doi: 10.1016/j.pneurobio.2015.09.001

Pezzulo, G., Rigoli, F. et Friston, K. (2018). Hierarchical Active Inference: A Theory of Motivated Control. *Trends in Cognitive Sciences*, 22(4), 294-306. doi: 10.1016/j.tics.2018.01.009

Pfaff, D. W. (1999). *Drive: Neurobiological and Molecular Mechanisms of Sexual*

Motivation. (s. l.) : MIT Press.

Phelps, E. A., Delgado, M. R., Nearing, K. I. et LeDoux, J. E. (2004). Extinction Learning in Humans: Role of the Amygdala and vmPFC. *Neuron*, 43(6), 897-905. doi: 10.1016/j.neuron.2004.08.042

Phillips, M. L., Drevets, W. C., Rauch, S. L. et Lane, R. (2003). Neurobiology of emotion perception I: the neural basis of normal emotion perception. *Biological Psychiatry*, 54(5), 504-514. doi: 10.1016/S0006-3223(03)00168-9

Pietrini, P., Guazzelli, M., Basso, G., Jaffe, K. et Grafman, J. (2000). Neural Correlates of Imaginal Aggressive Behavior Assessed by Positron Emission Tomography in Healthy Subjects. *American Journal of Psychiatry*, 157(11), 1772-1781. doi: 10.1176/appi.ajp.157.11.1772

Plebe, A. et Mazzone, M. (2016). Neural plasticity and concepts ontogeny. *Synthese*, 193(12), 3889-3929. doi: 10.1007/s11229-016-1131-z

Povinelli, D. J. (1995). The Unduplicated Self. Dans P. Rochat (dir.), *The Self in Infancy: Theory and Research* (vol. 112, p. 161-192). Elsevier. doi: 10.1016/S0166-4115(05)80011-1

Povinelli, D. J. et Simon, B. B. (1998). Young children's understanding of briefly versus extremely delayed images of the self: Emergence of the autobiographical stance. *Developmental Psychology*, 34(1), 188-194. doi: 10.1037/0012-1649.34.1.188

Power, M. et Brewin, C. R. (1991). From Freud to cognitive science: A contemporary account of the unconscious. *British Journal of Clinical Psychology*, 30(4), 289-310. doi: 10.1111/j.2044-8260.1991.tb00951.x

Qin, P. et Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *NeuroImage*, 57(3), 1221-1233. doi: 10.1016/j.neuroimage.2011.05.028

Raichle, M. E. (2015). The Brain's Default Mode Network. *Annual Review of Neuroscience*, 38(1), 433-447. doi: 10.1146/annurev-neuro-071013-014030

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A. et Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(2), 676-682. doi: 10.1073/pnas.98.2.676

Raichle, M. E. et Snyder, A. Z. (2007). A default mode of brain function: A brief history of an evolving idea. *NeuroImage*, 37(4), 1083-1090. doi: 10.1016/j.neuroimage.2007.02.041

Ramstead, M. J. D., Badcock, P. B. et Friston, K. (2017). Answering Schrödinger's question: A free-energy formulation. *Physics of Life Reviews*. doi: 10.1016/j.plrev.2017.09.001

Ramstead, M. J., Kirchhoff, M. D. et Friston, K. J. (2020). A tale of two densities: active inference is enactive inference. *Adaptive Behavior*, 28(4), 225-239. doi: 10.1177/1059712319862774

Ramstead, M. J., Veissière, S. P. et Kirmayer, L. J. (2016). Cultural affordances: scaffolding local worlds through shared intentionality and regimes of attention. *Frontiers in psychology*, 7, 1090. doi: 10.3389/fpsyg.2016.01090

Rangel, A., Camerer, C. et Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9(7),

545-556. doi: 10.1038/nrn2357

Rapaport, D. (1967). *Collected Papers*. New York; London : Basic Books.

Reinders, A. A. T. S., Nijenhuis, E. R. S., Paans, A. M. J., Korf, J., Willemsen, A. T. M. et den Boer, J. A. (2003). One brain, two selves. *NeuroImage*, 20(4), 2119-2125. doi: 10.1016/j.neuroimage.2003.08.021

Reinders, A. A. T. S., Nijenhuis, E. R. S., Quak, J., Korf, J., Haaksma, J., Paans, A. M. J., ... den Boer, J. A. (2006). Psychobiological Characteristics of Dissociative Identity Disorder: A Symptom Provocation Study. *Biological Psychiatry*, 60(7), 730-740. doi: 10.1016/j.biopsych.2005.12.019

Richerson, P. J. et Boyd, R. (2005). *Not by genes alone: how culture transformed human evolution*. Chicago : University of Chicago Press.

Rietveld, E., Denys, D. et Van Westen, M. (2018). Ecological-Enactive Cognition as engaging with a field of relevant affordances. Dans L. de Bruin, A. Newen et S. Gallagher, *The Oxford Handbook of 4E Cognition*. Oxford : Oxford University Press.

Roberts, A. M. et Thorpe, S. K. S. (2014). Challenges to human uniqueness: bipedalism, birth and brains. *Journal of Zoology*, 292(4), 281-289. doi: 10.1111/jzo.12112

Rosenkranz, J. A. et Grace, A. A. (2002). Cellular Mechanisms of Infralimbic and Prelimbic Prefrontal Cortical Inhibition and Dopaminergic Modulation of Basolateral Amygdala Neurons In Vivo. *Journal of Neuroscience*, 22(1), 324-337. doi: 10.1523/JNEUROSCI.22-01-00324.2002

Ruffman, T., Taumoepeau, M. et Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, 30(1),

87-104. doi: 10.1111/j.2044-835X.2011.02045.x

Satpute, A. B. et Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, 1079(1), 86-97. doi: 10.1016/j.brainres.2006.01.005

Satpute, A. B. et Lindquist, K. A. (2019). The Default Mode Network's Role in Discrete Emotion. *Trends in Cognitive Sciences*, 23(10), 851-864. doi: 10.1016/j.tics.2019.07.003

Schechtman, M. (2011). The Narrative Self. Dans S. Gallagher (dir.), *The Oxford Handbook of the Self*. (s. l.) : Oxford University Press.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T. et Vogeley, K. (2013). Toward a second-person neuroscience 1. *Behavioral and Brain Sciences*, 36(4), 393-414. doi: 10.1017/S0140525X12000660

Schore, A. N. (2016). *Affect regulation and the origin of the self: the neurobiology of emotional development*. New York, NY : Psychology Press.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, 26(1), 113-125. doi: 10.1023/A:1003044231033

Schrodinger, E. (2012). *What is Life?: With Mind and Matter and Autobiographical Sketches*. (s. l.) : Cambridge University Press.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., ... Greicius, M. D. (2007). Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. *Journal of Neuroscience*, 27(9), 2349-2356. doi: 10.1523/JNEUROSCI.5587-06.2007

Seemann, A. (2011). *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*. (s. l.) : MIT Press.

Sestieri, C., Corbetta, M., Romani, G. L. et Shulman, G. L. (2011). Episodic Memory Retrieval, Parietal Cortex, and the Default Mode Network: Functional and Topographic Analyses. *Journal of Neuroscience*, 31(12), 4407-4420. doi: 10.1523/JNEUROSCI.3335-10.2011

Seth, A. et Friston, K. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160007. doi: 10.1098/rstb.2016.0007

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11), 565-573. doi: 10.1016/j.tics.2013.09.007

Seth, A. K. (2014). The cybernetic Bayesian brain. Dans *Open Mind*. (s. l.) : Open MIND. Frankfurt am Main: MIND Group.

Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *American Psychologist*, 65(2), 98-109. doi: 10.1037/a0018378

Shedler, J. (2018). Where Is the Evidence for “Evidence-Based” Therapy? *Psychiatric Clinics of North America*, 41(2), 319-329. doi: 10.1016/j.psc.2018.02.001

Sheline, Y. I., Barch, D. M., Price, J. L., Rundle, M. M., Vaishnavi, S. N., Snyder, A. Z., ... Raichle, M. E. (2009). The default mode network and self-referential processes in depression. *Proceedings of the National Academy of Sciences*, 106(6), 1942-1947. doi: 10.1073/pnas.0812686106

Shin, L. M. (2006). Amygdala, Medial Prefrontal Cortex, and Hippocampal Function

in PTSD. *Annals of the New York Academy of Sciences*, 1071(1), 67-79. doi: 10.1196/annals.1364.007

Shin, L. M., Orr, S. P., Carson, M. A., Rauch, S. L., Macklin, M. L., Lasko, N. B., ... Pitman, R. K. (2004). Regional Cerebral Blood Flow in the Amygdala and Medial Prefrontal Cortex During Traumatic Imagery in Male and Female Vietnam Veterans With PTSD. *Archives of General Psychiatry*, 61(2), 168. doi: 10.1001/archpsyc.61.2.168

Simon, R. et Engström, M. (2015). The default mode network as a biomarker for monitoring the therapeutic effects of meditation. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.00776

Simons, J. S., Henson, R. N. A., Gilbert, S. J. et Fletcher, P. C. (2007). Separable Forms of Reality Monitoring Supported by Anterior Prefrontal Cortex. *Journal of Cognitive Neuroscience*, 20(3), 447-457. doi: 10.1162/jocn.2008.20036

Singer, J. A. et Blagov, P. (2004). The Integrative Function of Narrative Processing: Autobiographical Memory, Self-Defining Memories, and the Life Story of Identity. Dans *The self and memory* (p. 117-138). New York, NY, US : Psychology Press.

Singer, J. A. et Bluck, S. (2001). New Perspectives on Autobiographical Memory: The Integration of Narrative Processing and Autobiographical Reasoning. *Review of General Psychology*, 5(2), 91-99. doi: 10.1037/1089-2680.5.2.91

Siposan, D. G. et Aliu, O. F. (2014). Adjusting the light in the limit conditions of consciousness by the means of ascending reticular activating system (ARAS*) and of subordinated systems Part 1: Introduction and Aims. *LASER THERAPY*, 23(1), 31-40. doi: 10.5978/islsm.14-OR-04

Sleigh, M. (2011). Altricial. Dans S. Goldstein et J. A. Naglieri (dir.), *Encyclopedia of Child Behavior and Development* (p. 80-80). Boston, MA : Springer US. doi: 10.1007/978-0-387-79061-9_100

Smigielski, L., Scheidegger, M., Kometer, M. et Vollenweider, F. X. (2019). Psilocybin-assisted mindfulness training modulates self-consciousness and brain default mode network connectivity with lasting effects. *NeuroImage*, 196, 207-215. doi: 10.1016/j.neuroimage.2019.04.009

Solms, M. (2013). The Conscious Id. *Neuropsychanalysis*, 15(1), 5-19. doi: 10.1080/15294145.2013.10773711

Solms, M. et Friston, K. (2018). How and Why Consciousness Arises: Some Considerations From Physics and Physiology. *Journal of Consciousness Studies*, 25(5-6), 202–238.

Solms, M. L. (2018). The Neurobiological Underpinnings of Psychoanalytic Theory and Therapy. *Frontiers in Behavioral Neuroscience*, 12. doi: 10.3389/fnbeh.2018.00294

Solms, M. et Turnbull, O. (2018). *The Brain and the Inner World: An Introduction to the Neuroscience of Subjective Experience*. (s. l.) : Routledge.

Solomon, J. et George, C. (1996). Defining the caregiving system: Toward a theory of caregiving. *Infant Mental Health Journal*, 17(3), 183-197. doi: 10.1002/(SICI)1097-0355(199623)17:3<183::AID-IMHJ1>3.0.CO;2-Q

Somel, M., Franz, H., Yan, Z., Lorenc, A., Guo, S., Giger, T., ... Khaitovich, P. (2009). Transcriptional neoteny in the human brain. *Proceedings of the National Academy of*

Sciences, 106(14), 5743-5748. doi: 10.1073/pnas.0900544106

Sperber, D., Mascaro, O., Mercier, H., Origgi, G. et Wilson, D. (2010). *Epistemic Vigilance*, 35. doi: 10.1111/j.1468-0017.2010.01394.x

Spreng, Mar, R. A. et Kim, A. S. N. (2008). The Common Neural Basis of Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the Default Mode: A Quantitative Meta-analysis. *Journal of Cognitive Neuroscience*, 21(3), 489-510. doi: 10.1162/jocn.2008.21029

Spreng, R. N. et Andrews-Hanna, J. R. (2015). The Default Network and Social Cognition. Dans *Brain Mapping* (p. 165-169). Elsevier. doi: 10.1016/B978-0-12-397025-1.00173-1

Spreng, R. N., Stevens, W. D., Chamberlain, J. P., Gilmore, A. W. et Schacter, D. L. (2010). Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *NeuroImage*, 53(1), 303-317. doi: 10.1016/j.neuroimage.2010.06.016

Spunt, R. P., Meyer, M. L. et Lieberman, M. D. (2015). The Default Mode of Human Brain Function Primes the Intentional Stance. *Journal of Cognitive Neuroscience*, 27(6), 1116-1124. doi: 10.1162/jocn_a_00785

Spunt, R. P., Satpute, A. B. et Lieberman, M. D. (2010). Identifying the What, Why, and How of an Observed Action: An fMRI Study of Mentalizing and Mechanizing during Action Observation. *Journal of Cognitive Neuroscience*, 23(1), 63-74. doi: 10.1162/jocn.2010.21446

Sridharan, D., Levitin, D. J. et Menon, V. (2008). A critical role for the right fronto-

insular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences*, 105(34), 12569-12574. doi: 10.1073/pnas.0800005105

Steinert, C., Munder, T., Rabung, S., Hoyer, J. et Leichsenring, F. (2017). Psychodynamic Therapy: As Efficacious as Other Empirically Supported Treatments? A Meta-Analysis Testing Equivalence of Outcomes. *American Journal of Psychiatry*, 174(10), 943-953. doi: 10.1176/appi.ajp.2017.17010057

Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior*, 106(1), 5-15. doi: 10.1016/j.physbeh.2011.06.004

Stern, D. N. (1998). *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*. (s. 1.) : Karnac Books.

Stern, R. (1987). *Theories of the unconscious and theories of the self* (p. xiii, 282). Hillsdale, NJ, US : Analytic Press, Inc.

Tabibnia, G., Lieberman, M. D. et Craske, M. G. (2008). The lasting effect of words on feelings: Words may facilitate exposure effects to threatening images. *Emotion*, 8(3), 307-317. doi: 10.1037/1528-3542.8.3.307

Taubner, S., Kessler, H., Buchheim, A., Kächele, H. et Staun, L. (2011). The Role of Mentalization in the Psychoanalytic Treatment of Chronic Depression. *Psychiatry: Interpersonal and Biological Processes*, 74(1), 49-57. doi: 10.1521/psyc.2011.74.1.49

Terman, D. H. et Izhikevich, E. M. (2008). State space. *Scholarpedia*, 3(3), 1924. doi: 10.4249/scholarpedia.1924

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D.,

Hollinshead, M., ... Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3), 1125-1165. doi: 10.1152/jn.00338.2011

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. (s. l.) : Harvard University Press.

Tiet, Q. Q., Rosen, C., Cavella, S., Moos, R. H., Finney, J. W. et Yesavage, J. (2006). Coping, symptoms, and functioning outcomes of patients with posttraumatic stress disorder. *Journal of Traumatic Stress*, 19(6), 799-811. doi: 10.1002/jts.20185

Tognoli, E. et Kelso, J. A. S. (2014). The Metastable Brain. *Neuron*, 81(1), 35-48. doi: 10.1016/j.neuron.2013.12.022

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, Mass : Harvard University Press.

Tomasello, M. (2010). *Origins of Human Communication*. (s. l.) : MIT Press.

Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, Massachusetts ; London, England : Harvard University Press.

Tomasello, M. (2019). *Becoming Human: A Theory of Ontogeny*. Cambridge, Massachusetts : Belknap Press: An Imprint of Harvard University Press.

Tomasello, M. et Carpenter, M. (2007). Shared intentionality. *Developmental Science*, 10(1), 121-125. doi: 10.1111/j.1467-7687.2007.00573.x

Tomasello, M. et Gonzalez-Cabrera, I. (2017). The Role of Ontogeny in the Evolution of Human Cooperation. *Human Nature*, 28(3), 274-288. doi: 10.1007/s12110-017-

9291-1

Tsuda, I. (2009). Hypotheses on the functional roles of chaotic transitory dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19(1), 015113. doi: 10.1063/1.3076393

Uddin, L. Q., Iacoboni, M., Lange, C. et Keenan, J. P. (2007a). The self and social cognition: the role of cortical midline structures and mirror neurons. *Trends in Cognitive Sciences*, 11(4), 153-157. doi: 10.1016/j.tics.2007.01.001

Uddin, L. Q., Iacoboni, M., Lange, C. et Keenan, J. P. (2007b). The self and social cognition: the role of cortical midline structures and mirror neurons. *Trends in Cognitive Sciences*, 11(4), 153-157. doi: 10.1016/j.tics.2007.01.001

Vaillant, G. E. (1992). *Ego mechanisms of defense: A guide for clinicians and researchers* (p. xiv, 306). Arlington, VA, US : American Psychiatric Association.

Valk, S. L., Bernhardt, B. C., Böckler, A., Kanske, P. et Singer, T. (2016). Substrates of metacognition on perception and metacognition on higher-order cognition relate to different subsystems of the mentalizing network. *Human Brain Mapping*, 37(10), 3388-3399. doi: 10.1002/hbm.23247

van IJzendoorn, M. H. (1995). Adult attachment representations, parental responsiveness, and infant attachment: A meta-analysis on the predictive validity of the Adult Attachment Interview. *Psychological Bulletin*, 117(3), 387-403. doi: 10.1037/0033-2909.117.3.387

Van Overwalle, F. et Vandekerckhove, M. (2013). Implicit and explicit social mentalizing: dual processes driven by a shared neural network. *Frontiers in Human*

Neuroscience, 7. doi: 10.3389/fnhum.2013.00560

Vannini, P., O'Brien, J., O'Keefe, K., Pihlajamäki, M., LaViolette, P. et Sperling, R. A. (2011). What Goes Down Must Come Up: Role of the Posteromedial Cortices in Encoding and Retrieval. *Cerebral Cortex*, 21(1), 22-34. doi: 10.1093/cercor/bhq051

Varela, F. J., Thompson, E. et Rosch, E. (2000). *The embodied mind: cognitive science and human experience* (8. print). Cambridge, Mass. : MIT Press.

Vartanian, O., Bristol, A. S. et Kaufman, J. C. (2013). *Neuroscience of Creativity*. (s. l.) : MIT Press.

Vasil, J., Badcock, P. B., Constant, A., Friston, K. et Ramstead, M. J. D. (2020). A World Unto Itself: Human Communication as Active Inference. *Frontiers in Psychology*, 11. doi: 10.3389/fpsyg.2020.00417

Vatansever, D., Manktelow, A., Sahakian, B. J., Menon, D. K. et Stamatakis, E. A. (2018). Default Mode Network Engagement Beyond Self-Referential Internal Mentation. *Brain Connectivity*, 8(4), 245-253. doi: 10.1089/brain.2017.0489

Vatansever, D., Menon, D. K. et Stamatakis, E. A. (2017). Default mode contributions to automated information processing. *Proceedings of the National Academy of Sciences*, 114(48), 12821-12826. doi: 10.1073/pnas.1710521114

Waelder, R. (1960). *Basic theory of psychoanalysis* (p. xiii, 273). Oxford, England : International Univer. Press.

Wallerstein, R. S. (1989). The Psychotherapy Research Project of the Menninger Foundation: An overview. *Journal of Consulting and Clinical Psychology*, 57(2), 195-205. doi: 10.1037/0022-006X.57.2.195

Walonick, D. S. (1993). *General Systems Theory*. Récupéré de <https://www.statpac.org/walonick/systems-theory.htm>

Wampold. (1997). Methodological Problems in Identifying Efficacious Psychotherapies. *Psychotherapy Research*, 7(1), 21-43. doi: 10.1080/10503309712331331853

Wampold. (2007). Psychotherapy: The humanistic (and effective) treatment. *American Psychologist*, 62(8), 857-873. doi: 10.1037/0003-066X.62.8.857

Wampold. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, 14(3), 270-277. doi: 10.1002/wps.20238

Wampold et Imel, Z. E. (2015). *The great psychotherapy debate: the evidence for what makes psychotherapy work* (Second edition). New York : Routledge.

Wampold, Minami, T., Baskin, T. W. et Callen Tierney, S. (2002). A meta-(re)analysis of the effects of cognitive therapy versus 'other therapies' for depression. *Journal of Affective Disorders*, 68(2-3), 159-165. doi: 10.1016/S0165-0327(00)00287-1

Wampold, Mondin, G. W., Moody, M., Stich, F., Benson, K. et Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, « all must have prizes. » *Psychological Bulletin*, 122(3), 203-215. doi: 10.1037/0033-2909.122.3.203

Waxman, S. R. et Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, 13(6), 258-263. doi: 10.1016/j.tics.2009.03.006

Waxman, S. R. et Gelman, S. A. (2010). Different kinds of concepts and different kinds

of words: What words do for human cognition. Dans *The making of human concepts* (p. 99-129). New York, NY, US : Oxford University Press. doi: 10.1093/acprof:oso/9780199549221.003.06

Waxman, S. R. et Markow, D. B. (1995). Words as Invitations to Form Categories: Evidence from 12- to 13-Month-Old Infants. *Cognitive Psychology*, 29(3), 257-302. doi: 10.1006/cogp.1995.1016

Wiebking, C., de Greck, M., Duncan, N. W., Heinzl, A., Tempelmann, C. et Northoff, G. (2011). Are emotions associated with activity during rest or interoception? An exploratory fMRI study in healthy subjects. *Neuroscience Letters*, 491(1), 87-92. doi: 10.1016/j.neulet.2011.01.012

Wilson, G. T., Wilfley, D. E., Agras, W. S. et Bryson, S. W. (2011). Allegiance Bias and Therapist Effects: Results of a Randomized Controlled Trial of Binge Eating Disorder. *Clinical psychology : a publication of the Division of Clinical Psychology of the American Psychological Association*, 18(2), 119-125. doi: 10.1111/j.1468-2850.2011.01243.x

Winnicott, D. W. (2005). *Playing and Reality*. (s. l.) : Psychology Press.

Wolpe, N., Ingram, J. N., Tsvetanov, K. A., Geerligs, L., Kievit, R. A., Henson, R. N., ... Rowe, J. B. (2016). Ageing increases reliance on sensorimotor prediction through structural and functional differences in frontostriatal circuits. *Nature Communications*, 7(1), 13034. doi: 10.1038/ncomms13034

Wright, J. S. et Panksepp, J. (2012). An Evolutionary Framework to Understand Foraging, Wanting, and Desire: The Neuropsychology of the SEEKING System. *Neuropsychoanalysis*, 14(1), 5-39. doi: 10.1080/15294145.2012.10773683

Xie, X., Mulej Bratec, S., Schmid, G., Meng, C., Doll, A., Wohlschläger, A., ... Sorg, C. (2016). How do you make me feel better? Social cognitive emotion regulation and the default mode network. *NeuroImage*, 134, 270-280. doi: 10.1016/j.neuroimage.2016.04.015

Yan, C.-G., Chen, X., Li, L., Castellanos, F. X., Bai, T.-J., Bo, Q.-J., ... Zang, Y.-F. (2019). Reduced default mode network functional connectivity in patients with recurrent major depressive disorder. *Proceedings of the National Academy of Sciences*, 116(18), 9078-9083. doi: 10.1073/pnas.1900390116

Yang, X., Hu, L., Zeng, J., Tan, Y. et Cheng, B. (2016). Default mode network and frontolimbic gray matter abnormalities in patients with borderline personality disorder: A voxel-based meta-analysis. *Scientific Reports*, 6(1), 34247. doi: 10.1038/srep34247

Zahavi, D. (2008). *Subjectivity and Selfhood: Investigating the First-Person Perspective*. (s. l.) : MIT Press.

Zeidner, M. et Endler, N. S. (dir.). (1996). *Handbook of coping: theory, research, applications*. New York : Wiley.

Zeig, J. K. et Munion, W. M. (dir.). (1990). *What is psychotherapy? contemporary perspectives* (1st ed). San Francisco : Jossey-Bass Publishers.

Zeveloff, S. I. et Boyce, M. S. (1982). Why Human Neonates Are So Altricial. *The American Naturalist*, 120(4), 537-542. doi: 10.1086/284010

Zhou, Y., Friston, K., Zeidman, P., Chen, J., Li, S. et Razi, A. (2018). The Hierarchical Organization of the Default, Dorsal Attention and Salience Networks in Adolescents and Young Adults. *Cerebral Cortex*, 28(2), 726-737. doi: 10.1093/cercor/bhx307

