

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

REDÉFINITION NEUROCOGNITIVE DU PROBLÈME DE LA PERSISTANCE
DE L'IDENTITÉ PERSONNELLE À TRAVERS LE TEMPS

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAITRISE EN PHILOSOPHIE

PAR

SAFAE ESSAFI

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

Un immense merci à mon directeur de recherche, Professeur Pierre Poirier, pour son soutien, sa disponibilité hors du commun, sa flexibilité et son empathie.

Je tiens également à exprimer toute ma gratitude aux membres du jury, Professeur Luc Faucher et Professeur Alain Voizard, pour leur bienveillance et leur indulgence.

Un merci tout spécial à mon mari, collègue et complice, Simon Tremblay, pour sa présence essentielle et son soutien indéfectible. On n'a pas fini de faire des mauvais coups! Et merci à notre petit Alaric d'avoir attendu que je puisse faire le dépôt initial de ce mémoire avant de faire sa grande entrée en scène!

TABLE DES MATIÈRES

LISTE DES FIGURES.....	v
RÉSUMÉ	vi
INTRODUCTION	1
CHAPITRE I LE PROBLÈME DE LA PERSISTENCE DE L'IDENTITÉ PERSONNELLE À TRAVERS LE TEMPS.....	9
1.1 Deux explications intuitives s'opposent	11
1.1.1 Le caractère intuitif des deux positions et distinction entre identité numérique et identité qualitative.....	12
1.1.2 Limites atteintes : l'intuition comme fondement argumentatif donne lieu à un problème de circularité.....	28
CHAPITRE II REDÉFINITIONS NATURALISÉES DES CRITÈRES PSYCHOLOGIQUE ET CORPOREL DE LA PERSISTANCE DE L'IDENTITÉ PERSONNELLE À TRAVERS LE TEMPS.....	31
2.1 Le problème de la persistance de l'identité personnelle à travers le temps reformulé.....	38
2.1.1 Distinction entre nécessité phénoménologique et réalité ontologique....	40
2.1.2 Le soi hiérarchiquement construit : récupération des deux critères intuitifs	48
CHAPITRE III L'IDENTITÉ PERSONNELLE CONÇUE COMME LE RÉSULTAT DE LA MODÉLISATION HIÉRARCHIQUE DE SOI DANS LE MONDE	53
3.1 La hiérarchie du soi	54
3.1.1 Au-delà de la dichotomie corps-esprit	55
3.1.2 Modélisation de soi et du monde	64
CONCLUSION.....	75

BIBLIOGRAPHIE 79

LISTE DES FIGURES

Figure	Page
1.1 Critères de persistance de l'identité personnelle	23
1.2 Expériences de pensée	25
1.3 Problème de la circularité	29
2.1 L'esprit étendu	47
3.1 Le spectre du sentiment de réalité, de la transparence à l'opacité phénoménales	56
3.2 Division hiérarchique des niveaux du soi	59
3.3 Tirée de Metzinger, 2013b – énumération de niveaux hiérarchiques du soi	63
3.4 Boucle action-perception	67
3.5 Tirée de <i>The Ego Tunnel: the science of the mind and the myth of the self</i> (Metzinger, 2009)	68
3.6 Tirée de <i>Video Ergo Sum: Manipulating Bodily Self-Consciousness</i> (Lenggenhager et al., 2007).....	71

RÉSUMÉ

Les critères de persistance de l'identité personnelle à travers le temps sont souvent discutés sur la base d'expériences de pensée qui servent à sonder nos intuitions. Parmi les critères les plus abondamment discutés, on retrouve le critère psychologique et le critère corporel de l'identité personnelle. Nous soutenons toutefois que cette méthodologie est problématique puisqu'elle semble impliquer un raisonnement circulaire. En effet, le caractère intuitif d'un critère repose essentiellement sur notre phénoménologie de la persistance à travers le temps, c'est-à-dire sur notre impression subjective de persistance. Dans l'objectif d'offrir une solution au problème de la circularité, il est nécessaire de reformuler le problème de la persistance de l'identité personnelle à travers le temps sur des fondements naturalisés en faisant plein usage des ressources des neurosciences cognitives. Nous montrons qu'en tenant compte de la nature hiérarchique du soi l'opposition entre les critères psychologique et corporel n'a pas lieu d'être, la corporalité étant à la base de la hiérarchie du soi et la cognition étant un niveau hiérarchiquement supérieur. En tenant également compte de la nature inférentielle de l'identification hiérarchique de soi, nous montrons de plus que l'identification de soi est en fait relative et graduelle, relativisant du même coup le statut ontologique du soi en tant qu'entité immuable. À la lumière de tous ces éléments, il serait possible de réévaluer nos intuitions de façon à confirmer ou infirmer certains critères.

Mots clés : Identité personnelle, modélisation hiérarchique, théorie du traitement prédictif, intuition, naturalisme

INTRODUCTION

La question de la persistance de l'identité personnelle à travers le temps a toujours fasciné et continuera de faire partie de l'imaginaire populaire. Combien de scénarios de films ou de séries télévisées reposent sur ce thème. Pensons à ces classiques de la fin des années 80 et du début des années 90 tels que Robocop et Total Recall. L'intrigue de Robocop va comme suit : un policier tué dans le cadre de ses fonctions du nom de Murphy est transformé en cyborg. Il devient RoboCop : un policier mi-robot, mi-humain; un mélange de tissus organiques humains et de composantes robotiques. Bien que la transformation de Murphy en Robocop ait effacé sa mémoire, des souvenirs reviennent par bribes à ce dernier, qui commence à se souvenir d'événements de la vie de Murphy, dont sa mort. RoboCop apprend donc la vérité, soit qu'il est le résultat de la transformation de Murphy, et au fil du temps finit par s'auto-référencer comme étant Murphy, vivant en continuité avec ce que ce dernier était. Quant à Total Recall, il s'agit de l'histoire de Douglas Quaid, qui est troublé par un rêve récurrent se passant sur la planète Mars. On est en 2048 et Mars est colonisée par les Terriens. Quaid, en réponse à ce rêve, décide de faire appel à la compagnie Rekall, qui offre un service d'implantation de faux souvenirs. Au moment de l'implantation du faux souvenir de son souhait, un problème survient et Quaid se réveille en plein milieu de l'intervention, croyant être un agent secret poursuivi par des ennemis. Déstabilisé, l'employé chargé de l'implantation du faux souvenir ne comprend pas ce qui passe puisque l'implantation n'a pas encore eu lieu. Les responsables de l'implantation se chargent alors de lui administrer un sédatif et l'abandonnent, prenant soin d'effacer toute trace de son passage chez eux. En route vers chez lui, il est attaqué par des hommes lui reprochant d'être allé chez Rekall. Il les tue et continue sa route vers chez lui. Une fois

arrivé, sa femme, constatant ce qui vient de se produire, l'attaque également, mais finit par lui avouer qu'il n'est en fait pas Douglas Quaid. Il apprend alors que sa mémoire a été effacée et que ce qu'il croyait être son identité repose en fait sur des souvenirs qui lui ont été implantés. Il décide donc de se rendre sur Mars afin de découvrir sa « réelle » identité, celle de Hauser. Il découvrira que son ancienne et « véritable » identité était celle d'un homme mauvais, dont les valeurs sont radicalement opposées à celles qu'il a maintenant, avec sa nouvelle et « fausse » identité.

Ce qui caractérise ces deux histoires, c'est la tension entre l'identité psychologique et l'identité corporelle des personnages. Dans le premier cas, celui de RoboCop, le personnage subit une transformation corporelle importante et sa mémoire est effacée, mais une partie de son corps survit tout de même à la transformation et des souvenirs réapparaissent graduellement dans son esprit. Dans le deuxième cas, celui de Douglas Quaid, le personnage ne subit aucune transformation corporelle, mais sa mémoire est totalement effacée. Lorsqu'il découvre en quoi consistait son ancienne identité, il comprend que son identité actuelle est en contradiction totale avec cette dernière; ils n'ont rien en commun. Peut-on considérer qu'il s'agit de la même personne avant et après la transformation que chacun des personnages a subie? Est-ce que Murphy et RoboCop sont une seule et même personne? Certains diront que oui, car RoboCop possède les souvenirs de Murphy et que ses actions sont en continuité et en conformité avec les valeurs de Murphy. D'autres diront que non, car Murphy est mort, son corps ayant cessé d'exister, et que les souvenirs que RoboCop croit avoir sont en fait comparables à des événements imaginés. Est-ce que Quaid et Hauser sont une seule et même personne? Certains diront que oui, puisque peu importe qui il croit être, il s'agit effectivement du même individu, à l'image des personnes ayant subi d'importants dommages cérébraux entraînant une amnésie totale. D'autres diront que non, puisque Quaid n'a aucun souvenir de qui Hauser était, en plus d'être totalement opposé à celui-ci en termes de valeurs, croyances, désirs, etc.

Outre le fait que ces deux films soient hautement divertissants, ils ont l'avantage, en plus de fortement raviver la nostalgie des années 90, d'illustrer la tension entre deux critères hautement intuitifs de persistance de l'identité personnelle à travers le temps. Le premier fait appel au critère *psychologique* de la persistance de l'identité personnelle à travers le temps et le deuxième fait appel au critère *corporel* de la persistance de l'identité personnelle à travers le temps. Ce que de telles histoires arrivent à démontrer, c'est comment deux critères en apparence opposés, voire contradictoires, peuvent tous les deux être plausibles et intuitifs. En effet, les deux critères sont défendables, selon l'angle d'analyse utilisé.

Dans le premier chapitre de ce mémoire, nous analyserons ces deux critères en commençant par distinguer l'identité numérique de l'identité qualitative. Cette distinction permet d'expliquer dans quelles conditions nous privilégions un critère plutôt que l'autre et, par conséquent, pour quelle raison c'est tantôt l'un et tantôt l'autre qui nous semble intuitif. Nous distinguerons également les perspectives à la première et à la deuxième personne, qui sont respectivement subjective et intersubjective, de la perspective à la troisième personne, qui elle est objective. Nous arguerons qu'il n'est pas possible de déterminer rationnellement et de manière épistémologiquement valable quel est le critère de la persistance de l'identité personnelle à travers le temps si celui-ci est défendu sur la base d'une perspective à la première ou à la deuxième personne¹, donc fondé sur un raisonnement subjectif. L'impasse discursive qui en résulte est due à la circularité du raisonnement à l'origine des deux critères qui nous intéressent ici.

¹ Dans ce mémoire, les perspectives personnelles sont définies comme suit: la perspective à la première personne fait référence à une perspective subjective résultant de l'expérience du sujet; la perspective à la deuxième personne, quant à elle, fait référence à une perspective intersubjective, en ce qu'elle résulte de l'interaction entre deux sujets; finalement, la perspective à la troisième personne est une perspective objective car elle résulte de l'analyse d'un phénomène ou d'un objet donné indépendamment de l'expérience subjective de celui-ci.

En effet, autant le critère psychologique que le critère corporel de l'identité personnelle à travers le temps sont fondés sur des preuves subjectives, confirmées par des expériences de pensée. Cela donne lieu à un raisonnement fallacieux, en ce que la prémisse de celui-ci est contenue dans sa conclusion. Ce que nous disons lorsque nous fondons un critère de persistance de l'identité personnelle à travers le temps sur des expériences de pensée, donc sur des intuitions², c'est : « le critère de persistance de l'identité personnelle à travers le temps est x car j'ai l'impression (*intuition*) d'être la même personne à travers le temps sur la base de x , et inversement, j'ai l'impression (*intuition*) d'être la même personne à travers le temps sur la base de x car le critère de persistance de l'identité personnelle à travers le temps est x ». Évidemment, un tel raisonnement n'a aucune valeur objective et ne fait que confirmer nos propres intuitions.

Dans le deuxième chapitre de ce mémoire, nous rediscuterons ces deux critères en faisant appel à des notions qui ne reposent pas sur des expériences de pensée, dans une optique méthodologiquement naturaliste³. Nous montrerons de quelle manière le critère psychologique est un critère valable qui rend compte d'un aspect de notre identité personnelle en nous référant au travail de Joëlle Proust (2003), qui naturalise la théorie lockéenne de l'identité basée sur la mémoire et la fait échapper au piège de la circularité en posant de quelle manière l'existence de l'agent ne présuppose pas l'existence du soi. Ce processus de naturalisation des critères de l'identité personnelle

² Par intuitif, ou intuition, nous entendons ici « *feeling of correctness* », également appelé « *ascent feeling* », c'est-à-dire le sentiment (*feeling*) intellectuel d'avoir inféré une conclusion à partir de processus internes fiables (Wolos, 2016).

³ Nous utilisons ici le terme « naturaliste » pour désigner la méthode de recherche philosophique consistant à fonder les thèses sur des observations empiriques afin d'appuyer les intuitions philosophiques qui y ont donné naissance. Cette méthode ne nie pas l'importance des intuitions dans la recherche philosophique, mais les considère comme étant insuffisantes pour fonder des thèses de manière épistémologiquement valable. Voir les travaux de Hilary Kornblith (2002) ainsi que ceux de Joshua Knobe et Shaun Nichols (2008) pour plus de détails à ce sujet.

se poursuit avec la notion d'esprit étendu, qui montre de quelle manière l'identité ne se limite pas à la boîte crânienne (Clark & Chalmers, 1998) et que l'esprit, le corps et l'environnement sont dynamiquement liés, l'identité reposant en fait sur l'interaction des trois (Frith, 2007). Cette déconstruction de la rigidité catégorielle des critères de l'identité personnelle nous permettra d'ensuite introduire l'idée d'un soi⁴ hiérarchique, à la base duquel se trouve la conscience corporelle de soi. Nous nous référerons à la notion de soi pré-réflexif pour en faire la démonstration. L'élément central de notre argumentaire repose sur la distinction entre nécessité phénoménologique et réalité ontologique. Alors que l'identité personnelle est phénoménologiquement indéniable, en ce qu'il s'agit d'une nécessité phénoménologique, c'est-à-dire qu'il ne nous est pas possible de nier notre propre identité personnelle (à moins de souffrir d'une pathologie), la réalité de son existence, ontologiquement parlant, n'est quant à elle pas indéniable. Autrement dit, la nécessité phénoménologique d'une chose ne garantit pas sa nécessité métaphysique ou sa réalité ontologique. Nous appuierons nos propos sur les travaux de Thomas Metzinger (2003, 2010).

La récupération des deux critères en question, en démontrant leur complémentarité et en introduisant le caractère hiérarchique du soi, nous permettra de passer au second chapitre de ce mémoire. Il y sera question du rôle de l'identité personnelle en tant que processus inférentiel reposant sur une hiérarchie du soi, allant du niveau minimal corporel au niveau supérieur psychologique. Nous verrons de quelle manière nous en arrivons à « inférer » que notre corps *est* effectivement *notre* corps puis, de manière

⁴ Il est à noter que les termes « personne » et « soi » seront ici utilisés de manière interchangeable - bien que le terme « personne » soit théoriquement chargé, il ne sera ici utilisé que dans son sens le plus élémentaire et sera synonyme du terme « soi », tel qu'utilisé communément (*folk usage*). La notion de soi est explorée dans ce mémoire car il s'agit du concept au cœur de la question de la persistance de l'identité personnelle à travers le temps – ce qui persiste est la personne, ou, autrement dit, le soi.

générale, de quelle manière notre modèle de soi s'auto-représente, et devient capable d'organiser de façon métacognitive ses états futurs.

Dans ce mémoire, nous défendrons donc une définition hiérarchique et probabiliste de l'identité personnelle, accordant autant d'importance à l'aspect corporel qu'à l'aspect psychologique du soi. Dans ce cadre, l'opposition entre le critère psychologique et le critère corporel de la persistance de l'identité personnelle à travers le temps est remplacée par une intégration des deux positions, chacune faisant partie d'un continuum hiérarchique (en certains points interdépendants) dont l'interaction avec l'environnement est nécessaire à l'identité personnelle à travers le temps.

Pour en faire la démonstration, nous débuterons le troisième chapitre en introduisant de manière plus détaillée la classification hiérarchique du soi de Thomas Metzinger. Une meilleure compréhension des divers niveaux du soi nous permettra de concevoir de quelle manière l'aspect psychologique, ou cognitif, du soi est en fait indissociable du corps, sachant que la hiérarchie du soi tire son origine dans le soi minimal phénoménal, qui, lui, est ancré dans la corporalité. Nous verrons ainsi qu'il n'est pas pertinent d'opposer un critère corporel à un critère psychologique si l'on traite de la question de la persistance de l'identité personnelle à travers le temps sans d'abord distinguer les divers niveaux hiérarchiques du soi. En effet, le critère de persistance dépend du niveau hiérarchique visé par l'attribution d'identité.

Il sera ensuite temps d'introduire l'aspect inférentiel du soi, en présentant la théorie du traitement prédictif. La théorie du soi ici défendue, ainsi que la reformulation du problème de persistance de l'identité personnelle à travers le temps qui en découle, repose sur la théorie du traitement prédictif, qui elle, est fondée sur un principe

computationnel⁵ et dynamique⁶. Comme son nom l'indique, la théorie du traitement prédictif stipule que le cerveau est une « machine » à prédictions. Les estimations statistiques que celui-ci produit exigent la présence de représentations, qui elles servent à prédire les *inputs* sensoriels présents et futurs, via la constitution d'un modèle des causes de ceux-ci. Les erreurs de prédictions encourues, c'est-à-dire l'incongruence entre les *inputs* sensoriels prédits et réels, servent au calibrage de la « machine prédictive » à l'œuvre. Ce calibrage se fait via l'actualisation des représentations servants à la prédiction en question. Le but de cette opération d'actualisation est de minimiser le plus possible les erreurs de prédiction.

L'identité personnelle, comprise dans ce cadre explicatif, est donc le résultat de la représentation du corps et de l'esprit en tant qu'entités dans le modèle interne (de soi et du monde en général) de l'agent. Étant des « machines » à prédictions et à minimisation des erreurs de prédiction, nous inférons constamment les causes de la variété des *inputs* sensoriels (visuels, mais également proprioceptifs et intéroceptifs), jusqu'à créer des modèles du monde et de soi. C'est précisément par cette modélisation de soi (à commencer par son propre corps) que l'identité émerge. En attribuant certains *inputs* sensoriels récurrents à son propre corps, l'agent se représente au fil du temps toute une série de causes endogènes hiérarchiques et cachées qui finissent par constituer ce que nous pouvons appeler « soi ». La modélisation de soi par l'agent se fait via la minimisation des erreurs de prédiction, en tant que série hiérarchique de causes

⁵ Il est à noter qu'il existe des versions plus radicales de la théorie du traitement prédictif qui rejettent l'interprétation selon laquelle celle-ci est de nature computationnelle. Ce n'est pas le cas de la version ici présentée et soutenue. Celle-ci est basée sur les travaux de Jacob Hohwy (2013), Andy Clark (2016) et Wanja Wiese et Thomas Metzinger (2017).

⁶ Le terme dynamique doit être ici compris dans son sens le plus commun et le plus général. Il s'agit uniquement de décrire le caractère actif de l'actualisation des représentations.

endogènes cachées. C'est précisément ce qui donne lieu à l'identité personnelle et ce sont ces causes qui deviennent indissociables de soi. En effet, la capacité à constituer les causes endogènes en question repose sur la régularité et la récurrence des *inputs*. Ce principe s'appliquant à tous les niveaux hiérarchiques du soi, ce sont ces régularités et ces récurrences qui donneront lieu ultimement à *nos* désirs, croyances et à *notre* personnalité en général, ou autrement dit, à notre subjectivité. Nous pourrions alors revisiter et reconstruire en termes prédictivistes la définition de Proust de la constitution du soi explorée dans le premier chapitre.

Nous concluons donc ce mémoire en rappelant que la question de la persistance de l'identité personnelle à travers le temps n'a de sens que si l'on précise de quelle partie de l'identité personnelle il est question. Qu'est-ce qui persiste à travers le temps? De quel niveau hiérarchique est-il question? La question de la persistance de l'identité personnelle à travers le temps ne peut se poser de manière effective et épistémologiquement valable sans tenir compte de la définition de ce qu'est le soi, c'est-à-dire, selon la thèse défendue dans ce mémoire, une représentation hiérarchique probabiliste des causes endogènes des régularités de la variété de ses inputs sensoriels.

CHAPITRE I

LE PROBLÈME DE LA PERSISTENCE DE L'IDENTITÉ PERSONNELLE À TRAVERS LE TEMPS

Le problème de la persistance de l'identité personnelle à travers le temps peut être résumé en ces termes : une personne p est-elle la même à un temps t et à un temps t' ? Autrement dit, les changements subis par la personne p entre le temps t et le temps t' sont-ils tels qu'au temps t' , il ne s'agit plus de la personne p ? En d'autres mots, « if a person x exists at one time and a person y exists at another time, under what possible circumstances is it the case that x is y ? » (Olson, 2019) Pour répondre à ces questions, il faut avant tout déterminer quels sont les critères d'existence de la personne p . Ici, il n'est pas (uniquement) question de conditions corporelles de persistance, mais également de conditions psychologiques. C'est d'ailleurs précisément dans cette dichotomie que l'impasse quant à la résolution de la question trouve racine. Nous reviendrons sur ce point, qui est central à cette discussion, sous peu. De cette manière, il nous sera possible de déterminer ce qui constitue la personne p , et conséquemment, qu'est-ce qui ne constitue pas la personne p . Nous serons également en mesure de déterminer quels sont les changements pouvant mettre fin à l'existence de la personne p en tant que personne p .

Dans ce chapitre, nous commencerons par présenter les deux critères de l'identité personnelle opposés les plus populaires, soit le critère psychologique et le critère corporel, tels que définis par Derek Parfit (1984) et Judith Jarvis Thomson (2008). Plusieurs critères de la persistance de l'identité personnelle à travers le temps ont été

déendus dans le cadre de ce débat philosophique. On compte quatre catégories de critères au total : les positions psychologiques (Garrett, 1998; Hudson, 2001, 2007; Johnston, 1987, 2016; Lewis, 1976; Nagel, 1986; Noonan, 2003; Parfit, 1971, 1984, 2012; Perry, 1972; Shoemaker, 1970. 1984, 1997, 1999, 2008, 2011; Unger, 1990, 2000), les positions biologiques (Ayers, 1990; Carter, 1989; Mackie, 1999; Olson, 1997; van Inwagen, 1990; Williams, 1970), les positions narratives (Dennett, 1991; Schechtman, 1996; Schroer & Schroer, 2014) et, finalement, les positions « anti-circtérialistes » (Merricks, 1998; Swinburne, 1984; Lowe, 1996, 2012; Langford, 2017). Les plus populaires sont toutefois les deux premières catégories de critères, soit les positions psychologiques et les positions biologiques. Dans ce mémoire, nous avons choisi de nous concentrer sur deux articulations de ces deux positions, soit le critère de la continuité psychologique et le critère corporel de la persistance de l'identité personnelle à travers le temps. La popularité et le caractère intuitif de ces deux articulations sont les raisons qui ont motivé ce choix. Nous examinerons plus précisément leur caractère intuitif et de quelle manière ils correspondent respectivement à l'identité numérique et à l'identité qualitative. Nous discuterons de leur complémentarité, élément qui sera davantage développé dans les chapitres suivants de ce mémoire. Nous montrerons toutefois que la raison pour laquelle ils sont intuitifs et complémentaires est en fait la raison même pour laquelle ils ne peuvent fonder un critère de persistance de l'identité personnelle sans tomber dans le piège de la circularité. L'objection de la circularité a d'abord été développée par Joseph Butler en réponse à la thèse de Locke de la persistance de l'identité personnelle à travers les temps. Selon lui, la théorie de Locke est fallacieuse car il affirme que la conscience temporellement étendue est le critère de l'identité personnelle, alors que celle-ci présuppose l'existence d'une conscience temporellement étendue : « One should really think it self-evident that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity; any more than knowledge, in any other case, can constitute truth, which it presupposes. » (1842, p. 298) La démonstration du problème de circularité au cœur de la question discutée dans ce mémoire passera par une

distinction entre la nécessité phénoménologique et la nécessité métaphysique, telles que discutées par Thomas Metzinger (2010), ce qui nous permettra de distinguer la nécessité phénoménologique de la réalité ontologique de l'existence du soi en tant qu'entité. Cela nous mènera ultimement à présenter le caractère hiérarchique du soi, avec à sa racine la conscience pré-réflexive de soi, dont nous montrerons qu'elle est intrinsèquement corporelle et qui fonde le niveau plus élevé du soi qui lui est réflexif et donc psychologique. L'objectif de ce chapitre sera donc de montrer de quelle manière la phénoménologie du soi nous amène à percevoir, ou intuitionner, une opposition induite entre le critère psychologique et le critère corporel, mais qu'une fois la nécessité phénoménologique distinguée de la nécessité métaphysique du soi en tant qu'entité, nous pouvons constater que la réalité ontologique de la conscience de soi est en fait une combinaison interdépendante du corps et de l'esprit.

1.1 Deux explications intuitives s'opposent

Dans le débat de la persistance de l'identité personnelle à travers le temps, deux positions se démarquent par leur caractère particulièrement intuitif : l'approche de la continuité psychologique (Garrett, 1998; Hudson, 2001, 2007; Johnston, 1987, 2016; Lewis, 1976; Locke, 1689; Nagel, 1986; Noonan, 2003; Parfit, 1971, 1984, 2012; Perry, 1972; Shoemaker, 1970, 1984, 1997, 1999, 2008, 2011; Unger, 1990, 2000) et l'approche de la continuité corporelle (Ayer, 1936; Thomson, 2008; Williams, 1956, 1970). L'exercice est quasi infallible : lorsqu'on demande à un groupe de personnes (qu'il s'agisse de philosophes ou de non-philosophes) d'identifier quel est, selon eux, le critère de persistance de leur identité en tant qu'eux-mêmes, des caractéristiques psychologiques ou corporelles sont évoquées. Il est important de noter que, plus souvent qu'autrement, le critère psychologique est celui qui est d'emblée mentionné (Nichols & Bruno, 2010).

Dans l'ensemble, nous avons tendance à nous identifier à nos souvenirs, nos pensées et notre vie mentale. Bien que nous nous identifions également à nos attributs corporels, nous sommes d'ordinaire moins choqués par un changement corporel soudain que par un changement psychologique radical. Par exemple, pour la plupart, être le receveur d'un organe de porc dans le cadre d'une greffe ne signifie pas pour autant que nous cesserons d'être la même personne qu'avant la greffe en question; seule une petite composante de nous-même aura changé, comme ce serait le cas pour une voiture à laquelle on changerait les freins. On ne peut toutefois pas en dire autant pour les changements psychologiques radicaux ou soudains. Cette asymétrie se reflète même dans l'usage : on dit d'une personne qui souffre de démence ou de certains types d'amnésies qu'elle n'est « plus elle-même ». On n'utilise toutefois pas ce genre d'expressions pour faire référence à une personne ayant perdu une très grande quantité de poids en peu de temps. On dira d'elle qu'elle a subi une « transformation extrême », mais on n'aura pas tendance à se prononcer sur son identité personnelle.

1.1.1 Le caractère intuitif des deux positions et distinction entre identité numérique et identité qualitative

Dans la troisième partie de son livre *Reasons and Persons*, Derek Parfit (1984) expose sa conception de l'identité personnelle. Selon lui, rien ne nous permet de penser de manière rationnelle et justifiée que nous sommes effectivement la même personne au temps t et au temps t' . Pour expliquer son propos, Parfit fait appel à diverses expériences de pensée. Cela a pour objectif (et pour effet) de fonder sa thèse et ses arguments sur notre intuition. Selon l'expérience de pensée de la téléportation, à laquelle il a eu recours dans plusieurs de ses ouvrages, une machine à téléportation sur Terre a la capacité de détruire toutes les cellules de mon corps après avoir enregistré avec la plus grande exactitude leur disposition, information qui sera ensuite envoyée via ondes radio à une autre machine à téléportation, qui elle se trouve sur Mars et qui a la capacité de produire une réplique parfaite de mon corps à partir de matériel

biologique. La réplique qui se trouve sur Mars possède tous les souvenirs de la « version originale » qui se trouvait sur Terre. Autrement dit, autant son matériel corporel que son matériel mental sont en tous points identiques à ceux qui se trouvaient au point de départ sur Terre. La question qui se pose alors est « s'agit-il de la même personne » sur Terre et sur Mars? La personne qui était au départ sur Terre est-elle morte ou a-t-elle survécu sur Mars? Pour répondre à ces questions, Parfit soutient qu'il faut d'abord distinguer deux types d'identité : l'identité numérique et l'identité qualitative (Parfit 1984, p. 472). Deux choses peuvent être distinctes tout en étant similaires, en ce qu'il peut s'agir de deux objets tout en ayant en commun les mêmes caractéristiques. On dira alors de ces choses qu'elles sont numériquement distinctes (ce sont deux choses) qui sont qualitativement identiques (l'identité qualitative admettant des degrés). Par exemple, deux chaises peuvent être identiques, c'est-à-dire qu'il s'agit bel et bien de deux chaises (numériquement) distinctes, mais elles sont (qualitativement) identiques de par leurs caractéristiques telles que leurs couleurs, leurs matériaux, la disposition de ceux-ci, etc. Il en va de même pour les personnes : une seule et même personne peut en devenir une autre (qualitativement) et, vice versa, deux personnes (numériquement) distinctes peuvent être (qualitativement) identiques – par exemple, deux jumeaux peuvent être (qualitativement) identiques, physiquement parlant, même s'il s'agit de deux personnes (numériquement) distinctes.

There are two kinds of sameness, or identity. I and my Replica are qualitatively identical, or exactly alike. But we may not be numerically identical, or one and the same person. Similarly, two white billiard balls are not numerically but may be qualitatively identical. If I paint one of these balls red, it will cease to be qualitatively identical with itself as it was. But the red ball that I later see and the white ball that I painted red are numerically identical. They are one and the same ball. (Parfit, 1984, p. 472)

Parfit nous fait remarquer que c'est en général notre identité numérique qui nous préoccupe, en ce que celle-ci est indicatrice de notre survie :

When we are concerned about our future, it is our numerical identity that we are concerned about. I may believe that, after my marriage, I shall not be the same person. But this does not make marriage death. However much I change, I shall still be alive if there will be some person living who will be me. (Parfit, 1984, p. 473)

Autrement dit, l'identité numérique fait référence à notre identité (comprise en termes de survie) en tant qu'entité corporelle unique (ipséité⁷), alors que l'identité qualitative, elle, fait référence à notre identité en tant qu'entité mentale qui est la même à travers le temps (mêmeté). Être préoccupé par l'aspect numérique plutôt que qualitatif de l'identité, ou vice versa, dépend donc du critère de persistance auquel nous accordons de l'importance. L'idée, essentiellement, est d'identifier les critères nécessaires et suffisants pour la persistance de l'identité personnelle à travers le temps. Si c'est à l'identité numérique que l'on s'attache, alors c'est la continuité corporelle qui nous importe – une chose, bien que ses composantes puissent changer au fil du temps, demeure la même numériquement, car les changements en question suivent un parcours continu. Autrement dit, il n'y a jamais eu d'interruption dans la constitution de la chose en question. Ce critère est fortement intuitif. Si une personne perd un membre, tel un bras ou une jambe, et que celui-ci est remplacé par une prothèse, nous continuons de la percevoir comme étant la même personne qu'avant la perte du membre, puisqu'une continuité lie les deux moments charnières et que la perte ou le remplacement en question ne concerne pas toutes les parties du corps.

⁷ Nous utilisons le terme « ipséité » dans son sens ricœurien, c'est-à-dire pour faire référence à l'unicité d'un individu donné, ce qui en fait un individu totalement distinct d'un autre (Ricoeur, 1990).

Those who believe in the Physical Criterion would reject Teletransportation. They would believe this to be a way, not of travelling, but of dying. They would also reject, as inconceivable, reincarnation. They believe that someone cannot have a life after death, unless he lives this life in a resurrection of the very same, physically continuous body. (Parfit, 1984, p. 478)

La distinction entre identité numérique et identité qualitative s'illustre bien par la fameuse expérience de pensée du bateau de Thésée. Bien que la distinction en question ne soit ici aucunement liée à la question de l'identité personnelle, étant donné qu'il ne s'agit pas ici de l'identité d'une personne, mais plutôt de celle d'un objet inanimé, et qu'il ne soit par conséquent pas ici question des critères psychologique et corporel de l'identité personnelle, cette expérience de pensée permet tout de même de mieux saisir la distinction entre ces deux types d'identité. La légende veut que le bateau du héros Thésée, qui a dû partir combattre le minotaure, est resté au port pendant des années, de sorte que certaines de ses planches ont pourri et ont dû être une à une remplacées, jusqu'à toutes être remplacées au bout d'un certain nombre de décennies. S'agit-il alors toujours du bateau de Thésée? D'un côté, l'identité qualitative du bateau est intacte, en ce que les planches remplacées sont qualitativement les mêmes que celles d'origine. Mais de l'autre, son identité numérique est altérée, en ce qu'il n'est plus composé de ses pièces d'origine. Nous considérons intuitivement toutefois que le bateau qualitativement identique est toujours le bateau de Thésée, les changements ayant eu lieu ayant été graduels, respectant l'identité qualitative de l'objet et, surtout, sans briser la continuité physique de ce dernier. Autrement dit, selon ce critère d'identité à travers le temps, le bateau de Thésée est au temps t' le même que ce qu'il était au temps t , en ce qu'il est identique à lui-même qualitativement aux deux temps, n'ayant pas été reproduit ou détruit, et ce, bien que son identité numérique soit altérée. Il est important de préciser ici que l'identité numérique du bateau de Thésée est considérée comme étant altérée en raison de la catégorie conceptuelle qu'il représente et non en tant

qu'objet unique – son identité numérique est altérée car ses composantes d'origine ont été remplacées, altérant ainsi son « essence » numérique.

Cette expérience de pensée est utile pour mieux comprendre la distinction entre identité numérique et identité qualitative, mais, tel que précédemment mentionné, elle ne s'applique pas à l'identité personnelle. Il faut préciser qu'en ce qui concerne l'identité personnelle, toujours conformément à notre intuition, on ne semble pas accorder la même importance à toutes les parties du corps. En effet, la perte d'une jambe ou d'un bras n'équivaut pas, en termes de critère de persistance de l'identité personnelle, à la perte d'une partie du cerveau. Ce principe s'applique également au cerveau lui-même; il semble que nous n'accordons pas la même importance à toutes les parties cérébrales, selon leurs fonctions. Par exemple, une personne qui, suite à un accident, subit des dommages cérébraux affectant uniquement sa capacité à parler serait intuitivement toujours considérée comme étant la même qu'avant son accident. Par contre, dans le cas où les dommages causés engendrent un changement radical de la personnalité de l'individu en question, affectant ainsi son système de valeur, ses souvenirs, ses croyances, son caractère, etc., il deviendrait alors difficile de la qualifier comme étant la même qu'avant l'accident. Un exemple concret d'une telle situation est présenté dans le documentaire *A Totally Different Me*⁸, dans lequel Chris, après avoir subi un accident cérébral, est passé de jeune homme hétérosexuel fan de rugby à jeune coiffeur homosexuel passionné de décoration d'intérieur et de mode. Du moins, c'est ce que Chris prétend, mais très peu de gens semblent être de son avis. Chris est persuadé que son accident l'a « rendu homosexuel » et qu'il est aujourd'hui une personne totalement différente de ce qu'il était avant son accident, alors que d'autres pensent plutôt qu'il a toujours « été cette personne » sans le « réaliser ». Peu importe si Chris est réellement

⁸ <https://www.films.com/ecTitleDetail.aspx?TitleID=33332#>

devenu une nouvelle personne ou s'il a toujours été cette personne sans le réaliser, ce qui est intéressant dans cet exemple, c'est la géométrie variable des changements physiques : alors que certains changements, touchant davantage au corps ou à des parties cérébrales n'affectant pas ou peu la psychologie de l'individu, sont intuitivement considérés comme n'affectant pas l'identité à travers le temps, les changements physiques d'origine cérébrale et affectant la psychologie de l'individu sont plutôt intuitivement considérés comme affectant du même coup son identité à travers le temps. On semble intuitivement considérer qu'il y a coupure dans la continuité identitaire lorsqu'il y a changement psychologique radical (réel ou perçu) (et c'est ce qu'a vraisemblablement considéré le documentariste en intitulant son film « *A Totally Different Me* »). On voit donc que les critères psychologique et corporel de la persistance de l'identité personnelle à travers le temps sont intimement liés, puisque la continuité psychologique dépend d'une certaine continuité corporelle – mais ils sont néanmoins différents, en ce qu'ils ne sont pas traités de la même manière par notre intuition. Un autre exemple de changements psychologiques radicaux à considérer est celui de la stimulation cérébrale profonde. Cette intervention médicale est surtout utilisée pour traiter les patients atteints de la maladie de Parkinson, la dystonie, le trouble obsessionnel compulsif, ainsi que l'épilepsie, et elle consiste à implanter des électrodes dans le cerveau de manière à stimuler neurologiquement des régions cibles du cerveau. Toutefois, cette opération comporte parfois des effets secondaires indésirables donnant lieu à des changements importants dans la personnalité des patients traités. On peut entre autres observer l'apparition de comportements apathiques, d'hallucinations, d'hypersexualisation, de dysfonctions cognitives, de dépression et d'euphorie (Burn & Tröster, 2004).

Bien que ce soit notre identité numérique qui nous importe lorsqu'il s'agit de nous préoccuper de notre futur et/ou de notre survie, l'identité qualitative joue elle aussi un important rôle lorsqu'il est question de ces derniers. En effet, la continuité

psychologique peut également avoir une influence sur notre identité personnelle à travers le temps :

Though our chief concern is our numerical identity, psychological changes matter. Indeed, on one view, certain kinds of qualitative change destroy numerical identity. If certain things happen to me, the truth might not be that I become a very different person. The truth might be that I cease to exist – that the resulting person is someone else. (Parfit, 1984, p. 473)

Le critère psychologique de l'identité personnelle à travers le temps est défini par Parfit comme étant « the holding of overlapping chains of strong connectedness. » (Parfit, 1984, p. 482) Dans la suite d'événements mentaux dont est constituée notre vie mentale, chaque maillon doit être connecté à celui qui le précède et à celui qui le succède. Il y a continuité psychologique tant qu'il y a chevauchement des maillons de cette suite. C'est ce que Parfit appelle *Relation R*. Pour être la même personne à un temps t et à un temps t' , tous les maillons de la chaîne doivent se chevaucher de manière à ce qu'il y ait un lien de causalité entre chacun d'eux. C'est ce lien de causalité qui unit tous les événements mentaux que nous avons vécus depuis notre jeune âge jusqu'à aujourd'hui, et ce, sans que nous ayons besoin d'en garder le souvenir. Selon Parfit,

(1) There is psychological continuity if and only if there are overlapping chains of strong connectedness. X today is one and the same person as y at some past time if and only if (2) x is psychologically continuous with y , (3) this continuity has the right kind of cause, and (4) it has not taken

a "branching" form⁹. (5) Personal identity over time just consists in the holding of facts like (2) to (4). (Parfit, 1984, p. 484)

Bien que nous ne gardions pas de souvenirs précis de chaque événement de notre vie, il n'en demeure pas moins que tous ces événements sont interconnectés par une quelconque forme de vie psychologique, ce qui, par voie transitive, relie tous ces moments ensemble et nous permet, selon les défenseurs du critère psychologique de l'identité personnelle, d'affirmer que nous sommes la même personne au temps *t* et au temps *t'*.

Quant au critère corporel de l'identité personnelle, c'est vers Judith Jarvis Thomson (Thomson, 2008) que nous nous tournerons pour illustrer cette position. Tout comme pour le critère psychologique de l'identité personnelle, le critère corporel est intuitif de par sa simplicité et son évidence : nous reconnaissons les personnes qui nous entourent comme étant les mêmes à un temps *t* et à un temps *t'* sur la base de leur apparence physique et, de par notre nature incarnée, notre conception de nous-même est en tout temps indissociable de notre corps – nous expliquerons ce dernier point de manière plus précise dans les pages qui suivent. Comme le dit Thomson :

The simplest view of what people are is that they are their bodies. That view has other attractions besides its simplicity. I feel inclined to think that this fleshy object (my body is what I refer to) isn't something I merely currently inhabit: I feel inclined to think that it is me. This bony object (my left hand is what I refer to) – isn't it literally part of me? Certainly we all, at least at times, feel inclined to think that we are not

⁹ Parfit fait ici référence à la situation hypothétique, tout droit tirée d'expériences de pensée de duplication, où la relation de continuité se ramifierait, résultant ainsi en un dédoublement de la chaîne de connexité et mettant ainsi fin à l'ipséité de l'individu en question, en altérant son identité numérique.

merely embodied, but that we just, all simply, are our bodies. (Thomson, 2008, p. 155)

Selon celle-ci, la raison pour laquelle nous considérons le corps comme étant un critère intuitif de l'identité personnelle est que nous ne nous concevons pas comme étant une sorte de machine cartésienne, c'est-à-dire un esprit dans un corps. Au contraire, nous nous concevrions intuitivement, sans avoir besoin d'y penser, comme étant notre corps. Pour Thomson, une personne x est à un temps t la même qu'une personne y à un temps t' si et seulement si le corps de la personne x à t est le même que le corps de la personne y à t' , c'est-à-dire si le corps en question possède la même identité numérique aux deux temps. Autrement dit, le corps de x doit être au temps t en continuité avec celui de y au temps t' . Parfit caractérise la position de Thomson comme suit : « On Thomson's view, if her brain were destroyed, and my brain were successfully transplanted into her empty skull, it would be she who would wake up, mistakenly believing that she was me. » (Parfit, 2008, p. 178)

Thomson fait elle aussi appel aux expériences de pensée pour défendre sa position, ou plutôt pour démontrer en quoi le critère psychologique de l'identité personnelle ne satisfait pas aux exigences de l'identité personnelle pour être considéré comme tel. Elle nous fait entre autres remarquer notre rejet de l'indétermination – pourquoi refuse-t-on de suspendre notre jugement quant à la survie ou non du personnage d'expérience de pensée ayant subi une transplantation, une reprogrammation, une division ou une encore une téléportation? En effet, dans la plupart des expériences de pensée traitant d'identité personnelle, nous sentons le besoin de déterminer un survivant de l'opération de transformation, bien qu'il n'y ait souvent pas de réponse objectivement déterminable. Cela est dû au fait que ces expériences de pensée font appel à nos intuitions – il s'agit de « pompes à intuition » (Dennett, 1980). Or, de tels exercices étant fondés sur la manière dont nous percevons subjectivement notre propre identité et notre mode de fonctionnement cognitif étant tel qu'il ne permet pas l'indétermination,

puisque la cognition a pour but de résoudre l'incertitude ¹⁰, notre rejet de l'indétermination en réponse aux diverses expériences de pensée touchant à l'identité personnelle est inévitable. C'est ce que nous appellerons le piège de l'intuition dans le débat sur la persistance de l'identité personnelle à travers le temps, soit le raisonnement circulaire qui consiste à baser le critère d'identité sur notre conception de notre propre identité à travers le temps, donc sur notre intuition. Autant le critère psychologique que le critère corporel sont victimes de ce piège. Parfit (2008, p. 181) ainsi que Thomson (2008, p. 156) prétendent tous les deux que le critère qu'ils défendent respectivement est davantage conforme aux intuitions de la majorité, ce qui, selon eux, serait un incitatif à adopter le critère en question. Les expériences de pensée proposées par Parfit (2008) et Thomson (2008) pour montrer le bien-fondé de leurs critères respectifs s'inspirent de l'expérience de pensée de John Locke (1689), mais au lieu du prince et du savetier, c'est de Brown, Robinson et Dickenson qu'il s'agit. Elles se déclinent en cinq cas de figure que nous détaillerons sous peu. Voyons d'abord, dans ce premier schéma (figure 1.1), ci-dessous, les cinq critères selon lesquels on répond aux expériences de pensée de Parfit (2008) et Thomson (2008). Il s'agit, en premier lieu, du critère physique, ou corporel, selon lequel une personne est la même à un temps t (personne x) et à un temps t' (personne y) si et seulement si le corps de la personne à un temps t est le même que le corps de la personne à un temps t' . Il est en deuxième lieu question du critère psychologique, qui se décline en deux versions, la pure et l'impure. Selon la première version, soit la pure, une personne est la même à un temps t (personne x) et à un temps t' (personne y) si et seulement si la personne y est psychologiquement connectée à un temps t' à la personne x à un temps t . Selon la deuxième version, soit l'impure, une personne est la même à un temps t (personne x) et à un temps t' (personne y) si et seulement si la personne y est psychologiquement

¹⁰ Ce point sera explicité dans le deuxième chapitre de ce mémoire.

connectée à un temps t' à la personne x à un temps t et que la personne x et la personne y possèdent (numériquement parlant) le même cerveau, en plus d'être les seules qui soient psychologiquement connectées (clause de non compétition), c'est-à-dire que la connexité psychologique doit être uniquement entre la personne x et la personne y . Quant au critère hybride, il s'agit, comme son nom l'indique, de la combinaison du critère psychologique et du critère corporel : une personne x est la même à un temps t qu'une personne y à un temps t' si et seulement si le corps de la personne y est psychologiquement connecté à un temps t' à la personne x au temps t . Autrement dit, selon ce critère, pour qu'une personne demeure la même à travers le temps, il doit y avoir une continuité entre le corps existant à t' et la vie mentale (désirs, croyances, souvenirs, etc.) existant à t . Une variante de ce critère implique des « phases temporelles », c'est-à-dire des stades à diverses périodes d'une même vie, dont nous serions tous composés et qui feraient en sorte qu'il est possible d'être, à un stade donné, une seule et même personne puis, à un stade ultérieur, devenir deux personnes différentes. Un exemple fréquemment donné pour illustrer cette thèse est celui de deux routes qui convergent puis divergent : elles sont une seule et même route à un moment puis deviennent deux routes distinctes à un autre moment (Olson, 2021). Finalement, il y a le critère dit « ordinaire », c'est-à-dire le critère du sens commun, celui que nous utilisons d'emblée dans la vie de tous les jours, selon lequel une personne est la même à un temps t et à un temps t' si et seulement si la personne possède le même corps (numériquement parlant) aux deux temps. Ce critère est très similaire au critère corporel, à l'exception que le critère ordinaire ne permet pas l'indétermination de l'identité (c'est-à-dire l'indétermination quant à quelle est l'identité de la personne y à

un temps t'), alors que le critère corporel la postule dans certaines situations, comme nous le verrons dans le deuxième schéma ci-dessous (figure 1.2).

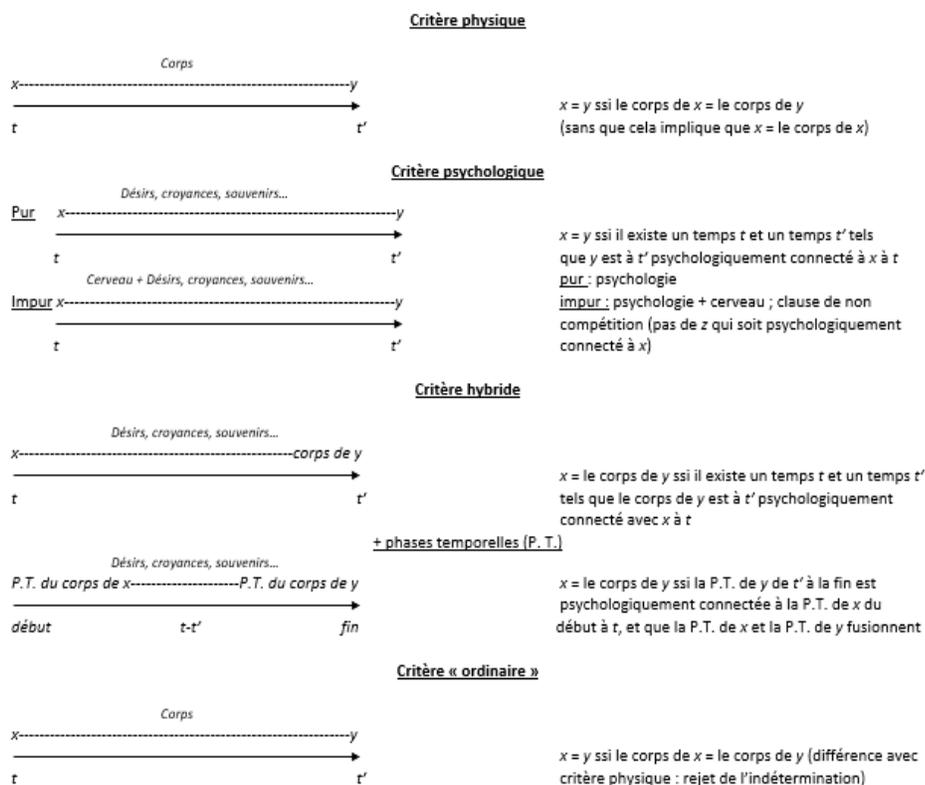


Figure 1.1 Critères de persistance de l'identité personnelle

Revenons maintenant aux expériences de pensée proposées par Parfit (2008) et Thomson (2008). Ce deuxième schéma (figure 1.2) montre la réponse à la question posée par chacun des scénarios en fonction du critère de persistance de l'identité personnelle présenté dans le schéma précédent (figure 1.1). Le premier scénario illustré dans la figure 1.2, *Brown-Case-One*, décrit une situation dans laquelle le cerveau de

Brown ainsi que sa vie mentale sont transplantés dans un autre corps, la question qui est posée est de savoir si Brown à un temps t et le survivant de la transplantation à un temps t' sont une seule et même personne. La réponse à cette question varie selon le critère de persistance choisi : selon le critère psychologique dit pur, c'est-à-dire celui qui exige la continuité de la vie mentale uniquement, ainsi que le critère psychologique impur, c'est-à-dire celui exigeant la continuité du cerveau en plus de la vie mentale, Brown est bien le survivant de la transplantation, ou, autrement dit, Brown et le survivant de l'opération sont bien la même personne. Toutefois, selon le critère corporel, le critère hybride ainsi que le critère dit ordinaire, Brown et le survivant ne sont pas la même personne, c'est-à-dire que Brown n'est pas le survivant de l'opération. Quant au deuxième scénario, *Brown-Case-Two*, le cerveau de Brown est transplanté dans un autre corps sans que sa vie mentale ne soit, elle aussi, transplantée. Dans ce scénario, Brown et le survivant sont deux personnes distinctes selon tous les critères. Le troisième scénario, *Brown-Case-Three*, décrit quant à lui une opération lors de laquelle seule la vie mentale de Brown est transplantée dans un autre corps. La survie de Brown n'est ici possible que si l'on adopte le critère psychologique pur ou le critère hybride en plus de l'existence de phases temporelles. Les deux derniers scénarios, *Brown-Case-Four* et *Brown-Case-Five*, impliquent deux survivants. Dans le *Brown-Case-Four*, les cerveaux de Robinson et de Dickenson sont reprogrammés avec la vie mentale de Brown. Dans ce scénario, tous les critères ont pour résultat la mort de Brown. Dans le *Brown-Case-Five*, le corps de Brown est divisé en deux parties qui deviendront chacune un corps à part entière. Dans ce scénario également, aucun critère ne permet d'affirmer que le survivant est Brown. Toutefois, le critère corporel, contrairement aux autres critères, ne permet tout simplement pas de se prononcer sur l'état de Brown : alors que les autres critères permettent de conclure que Brown ne survit pas à l'opération, le critère corporel permet l'indétermination comme réponse à la question de la survie de Brown.

Brown-Case-One :								
Transplantation du cerveau de Brown ainsi que de sa psychologie dans un autre corps			Critère psychologique pur	Critère psychologique impur	Critère physique	Critère hybride	Critère « ordinaire »	
<i>Cerveau + Désirs, croyances, souvenirs...</i>								
<i>Brown</i> ----- <i>Survivant</i>		Survivant=Brown	x	x				
t-----t'		Survivant≠Brown			x	x	x	
		Indéterminé						
Brown-Case-Two :								
Transplantation du cerveau de Brown, mais pas de sa psychologie, dans un autre corps			Critère psychologique pur	Critère psychologique impur	Critère physique	Critère hybride	Critère « ordinaire »	
<i>Cerveau</i>								
<i>Brown</i> ----- <i>Survivant</i>		Survivant=Brown						
t-----t'		Survivant≠Brown	x	x	x	x	x	
		Indéterminé						
Brown-Case-Three :								
Transplantation de la psychologie de Brown, mais pas de son cerveau, dans un autre corps			Critère psychologique pur	Critère psychologique impur	Critère physique	Critère hybride	Critère « ordinaire »	
<i>Désirs, croyances, souvenirs...</i>								
<i>Brown</i> ----- <i>Survivant</i>		Survivant=Brown	x			x (si P. T.)		
t-----t'		Survivant≠Brown		x	x		x	
		Indéterminé						
Brown-Case-Four:								
Reprogrammation du cerveau de Robinson ainsi que de Dickenson avec la psychologie de Brown			Critère psychologique pur	Critère psychologique impur	Critère physique	Critère hybride	Critère « ordinaire »	
<i>Désirs, croyances, souvenirs...</i>								
<i>Brown</i> ----- <i>Survivant1 + Survivant2</i>		Survivant=Brown						
t-----t'		Survivant≠Brown	x	x	x	x	x	
		Indéterminé						
Brown-Case-Five:								
Division du corps de Brown en deux parties, qui seront chacune reconstituées en deux corps			Critère psychologique pur	Critère psychologique impur	Critère physique	Critère hybride	Critère « ordinaire »	
<i>Corps</i>								
<i>Brown</i> ----- <i>Survivant1 + Survivant2</i>		Survivant=Brown						
t-----t'		Survivant≠Brown	x	x		x	x	
		Indéterminé			x			

Figure 1.2 Expériences de pensée

Intuitivement, nous n'envisageons même pas la possibilité que le survivant d'une expérience de pensée, d'une transformation extrême ou d'un accident ne soit pas déterminé. Thomson affirme qu'il nous faudrait combattre cette inclination et accepter l'indétermination quant à la survie. C'est selon elle la principale différence entre le critère corporel tel que défendu par les philosophes – qui accepte l'indétermination – et le critère dit « ordinaire », c'est-à-dire le critère du sens commun. Il est vrai que, de manière générale, il est impensable d'accepter une telle indétermination; dans la vie de

tous les jours, nous « savons » à qui nous avons affaire ou qui nous sommes¹¹. Même dans la majorité des cas de troubles mentaux affectant le contact à la réalité, tel que les cas de psychose ou de dépersonnalisation, une telle idée d'indétermination n'a pas lieu. Si Thomson est donc justifiée de distinguer ces deux types d'intuitions, elle ne s'extirpe toutefois pas du piège de l'intuition pour autant en défendant un critère corporel de l'identité personnelle. Elle se contente en fait d'affirmer un critère intuitif de l'identité personnelle du point de vue de la reconnaissance de soi-même et d'autrui. Dans les faits, rien ne nous justifie de soutenir davantage l'explication corporelle que l'explication psychologique, et vice versa. En fait, les deux explications semblent sensées, selon l'angle d'analyse utilisé. Elles sont même complémentaires.

Ainsi nous avons vu que le critère psychologique de la persistance de l'identité personnelle à travers le temps nous semble tout à fait évident et incontestable lorsqu'il constitue la réponse à la question : « Comment sais-je que je suis la même personne aujourd'hui que celle que j'étais hier ou même il y a 20 ans? » Ce critère devient toutefois bien plus critiquable et douteux lorsqu'on le propose comme réponse à la question : « comment sais-je que ma collègue est la même personne aujourd'hui que celle qu'elle était hier ou il y a 20 ans? » Inversement, il en va de même pour le critère corporel de la persistance de l'identité personnelle à travers le temps; il s'agit de la meilleure manière de reconnaître l'identité d'autrui, mais il n'est certainement pas gage de notre propre identité à travers le temps. Le critère corporel est ce que nous utilisons pour reconnaître les gens à travers le temps, et ce, même les personnes aux allures très semblables – nous distinguons les jumeaux identiques en nous référant à certaines

¹¹ Nous pourrions toutefois considérer que certains cas de changements radicaux de la personnalité résultent en une indétermination. Un exemple courant de ce type d'indétermination est celui des cas de verdicts de non-responsabilité criminelle pour cause de troubles mentaux ayant causé une suspension passagère et ponctuelle du jugement de la nature des actes commis.

caractéristiques physiques comme les grains de beauté. Le critère corporel est gage d'ipséité : il n'existe pas deux individus en tous points identiques sur la planète. Il y a bien sûr des personnes très similaires, des sosies, mais jamais identiques en tous points. Le critère corporel est donc garant de l'ipséité de l'individu et exprime l'identité numérique de toute chose. Il est également intuitif en raison de notre nature incarnée, tel que précédemment mentionné – nous sommes notre corps. Le critère corporel est donc fiable du point de vue de la première et de la deuxième personne en ce qui a trait à l'ipséité. Quant au critère psychologique de la persistance de l'identité personnelle, il nous indique surtout et principalement notre propre identité temporelle; il décrit ce qui est accessible à la première personne. À la deuxième et à la troisième personne, ce critère est plus difficilement fiable, car nous n'avons jamais réellement accès aux états mentaux d'autrui, nous avons uniquement accès à une description verbale de ceux-ci par la personne à qui ils appartiennent. Le critère psychologique est toutefois garant de notre mêmeté : il nous confirme notre propre identité à travers le temps par voie de continuité dans nos états mentaux. C'est précisément pour cette raison que le critère psychologique ne peut garantir la mêmeté d'autrui; les croyances, les désirs et autres états mentaux ne sont pas uniques lorsque rapportés verbalement – plusieurs personnes aiment la crème glacée au chocolat, il ne s'agit pas là d'une caractéristique psychologique me permettant de distinguer entre deux personnes, ni me permet-elle de confirmer l'identité d'une personne à travers le temps. Il en va de même pour le critère corporel en ce qui a trait à ma propre identité à travers le temps : le fait d'avoir le même corps hier et aujourd'hui ne me permet pas de confirmer que je suis la même aujourd'hui que celle que j'étais hier. Mon corps me prouve mon existence là et maintenant, en raison de ma nature incarnée, mais il ne m'informe pas sur mon identité à travers le temps. Je pourrais par exemple souffrir d'un trouble dissociatif, tel que le trouble dissociatif de l'identité, qui m'empêcherait de me reconnaître comme étant la même que j'étais en me fiant uniquement sur mon apparence physique.

1.1.2 Limites atteintes : l'intuition comme fondement argumentatif donne lieu à un problème de circularité

Nous soutenons que le fait de fonder des critères de l'identité personnelle sur des bases intuitives uniquement donne lieu à une impasse discursive. Chacun des critères de persistance de l'identité personnelle est posé comme constituant l'unique réponse à la question et comme étant opposé aux autres critères proposés. Ce que l'on constate toutefois, comme on a pu le voir dans les lignes qui précèdent, c'est que ces cadres conceptuels sont en fait complémentaires et capturent, chacun d'eux, un aspect distinct de la persistance de l'identité personnelle. Le critère psychologique est intuitif car il caractérise la perspective à la première personne de la persistance de l'identité personnelle. Le critère corporel est intuitif car il caractérise la perspective à la deuxième personne (résultant de l'interaction interpersonnelle) de la persistance de l'identité personnelle. Ces deux positions intuitives capturent par conséquent, chacune à leur manière, une facette de notre mode de fonctionnement cognitif. Cette situation est semblable à celle des concepts tels qu'Edouard Machery les comprend (Machery, 2009). En effet, trois théories se disputent depuis les années 1970 le titre de « la bonne théorie des concepts » : la théorie des prototypes, la théorie des exemplaires et la théorie des théories. Or Machery remarque qu'elles sont en fait complémentaires, chacune caractérisant une manière dont les corps d'information (*bodies of information*) sont encodés et utilisés par nos systèmes cognitifs.

Constater la complémentarité des critères psychologique et corporel, cependant, ne suffit pas à résoudre la question de la persistance de l'identité personnelle à travers le temps. En effet, l'erreur réside dans la manière dont la question est formulée : on tente de fournir une réponse objective, à savoir « quel est le critère d'identité à travers le temps », à partir d'observations à la première et à la deuxième personne – donc à partir d'une analyse subjective-mais celles-ci sont le fruit des mécanismes cognitifs mêmes nous permettant de nous considérer comme étant les mêmes à travers le temps et des

ressentis phénoménologiques générés par l'action de ces mécanismes. Les perspectives à la première et à la deuxième personne relèvent de notre vécu phénoménologique et ne peuvent d'aucune manière nous informer fidèlement sur la nature de notre identité. Le raisonnement devient aussitôt circulaire et relève de la pétition de principe : nous basons le critère de l'identité personnelle à travers le temps (corporel ou psychologique par exemple) sur nos observations subjectives de notre propre identité personnelle (sur la base d'expériences de pensée auxquelles on répond à partir de notre perspective à la première ou à la deuxième personne), de sorte que notre conclusion (critère d'identité) est contenue dans notre prémisse (notre propre impression d'identité à travers le temps).

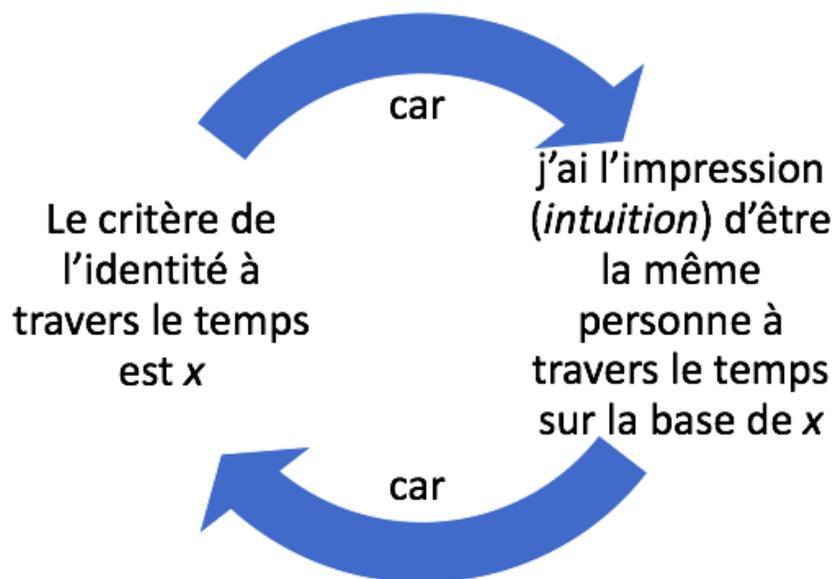


Figure 1.3 Problème de la circularité

Notre système perceptuel est un bon exemple de cette incongruence. Ce que nous voyons ne correspond pas exactement à la chose que nous sommes en train de percevoir. Notre système perceptuel reconstruit l'information sensorielle environnementale de

manière à pouvoir l'interpréter, c'est-à-dire la rendre traitable par les autres systèmes de la cognition (mémoire, prise de décision, action, etc.). Si nous expliquons un phénomène observé en nous référant directement et uniquement à ce que nous percevons, alors nous ne faisons rien d'autre que raisonner de manière circulaire. Par exemple, quand je dis « cet objet est rouge », au fond je dis « cet objet m'apparaît rouge ». Il s'en suit que je fais une explication circulaire si on me demande pourquoi l'objet est rouge et que je réponds que c'est parce qu'il m'apparaît rouge. De même, j'aurai commis une pétition de principe si je fais un raisonnement déductif dont la conclusion est « cet objet est rouge » et une prémisse est « cet objet m'apparaît rouge ». Il est évident que la seule conclusion que l'on puisse logiquement tirer d'une analyse subjective de l'objet en question dépendant de sa perception subjective, de sorte que la réponse à la question ne peut être objective. Nous reviendrons plus en détails sur le mécanisme cognitif à la base des jugements d'identité personnelle et des vécus phénoménologiques associés à ces jugements. Ce qu'il faut retenir à ce stade-ci, c'est que la nature, dans une certaine mesure (que nous expliquerons ci-dessous), illusoire de notre vécu phénoménologique nous empêche d'affirmer de manière épistémologiquement rigoureuse des faits à propos de la nature ontologique de notre identité à travers le temps.

Il faudrait d'abord comprendre les mécanismes cognitifs qui nous permettent de nous considérer nous-mêmes comme étant des personnes qui existent – car les critères que proposent aussi bien Parfit que Thomson pour affirmer qu'une personne sera ou non la même à un temps ultérieur les présupposent – pour ensuite être en mesure de comprendre le caractère intuitif des deux critères de l'identité personnelle. Il serait en effet absurde de tenter de déterminer quel est le critère de l'identité personnelle à travers le temps sans d'abord savoir ce qu'est une personne (soi).

Les tenants du critère corporel fondent ce dernier sur l'idée selon laquelle nous utilisons communément le corps comme critère de reconnaissance de soi et d'autrui. Selon eux,

nous ne nous demandons jamais si nous « faisons partie » de notre corps ou si telle ou telle partie de notre corps est bien la nôtre (à part dans des cas pathologiques). Au contraire, nous n'avons pas besoin d'y penser, nous sommes notre corps (Thomson, 2008, p. 155). Ils n'ont pas tort. De la même manière, les tenants du critère psychologique fondent ce dernier sur l'idée selon laquelle nous considérons de manière générale la continuité psychologique comme étant ce qui nous permet de nous considérer comme étant les mêmes que nous étions à un temps antérieur. Eux non plus n'ont pas tort. À moins de souffrir d'une pathologie qui nous empêche de nous auto-concevoir (c'est-à-dire avoir un concept de soi, grâce à un processus réflexif¹²)¹³, nous avons la capacité de consciemment avoir un effet sur nous-même. On peut supposer que c'est précisément sur cette capacité d'auto-affection – que Joëlle Proust définit comme étant « the general ability of taking new mental states and properties as goals of action, and of pursuing these goals » (2003, p. 500) – que se fondent les critères de l'identité personnelle établis par Parfit et Thomson; c'est cette capacité d'auto-affection qui rend intuitifs les critères psychologique et corporel de l'identité personnelle. Toutefois, tel que précédemment mentionné, le caractère intuitif d'un critère ne suffit pas pour en faire un critère valide. Pour déterminer la validité des critères à l'étude, il faut les naturaliser, c'est-à-dire les fonder sur des preuves empiriques et non simplement sur des intuitions.

¹² Nous distinguons dans ce mémoire la réflexivité de la réflectivité, conformément à la distinction tracée par Jerrold Seigel (2005). Selon ce dernier, le concept de réflexivité fait référence aux expériences s'apparentant au réflex (involontaires et passives), alors que celui de réflectivité fait plutôt référence à l'expérience volontaire (de second ordre).

¹³ La capacité de nous auto-concevoir est intrinsèque à toute personne puisque, tel que l'affirme Charles Taylor, l'humain est un animal qui s'auto-interprète (1985), c'est-à-dire que sa conception de lui-même est indissociable de son existence même.

CHAPITRE II

REDÉFINITIONS NATURALISÉES DES CRITÈRES PSYCHOLOGIQUE ET CORPOREL DE LA PERSISTANCE DE L'IDENTITÉ PERSONNELLE À TRAVERS LE TEMPS

Selon Proust, une condition nécessaire pour pouvoir être dite une personne consiste à être capable de métacognition : « to be a self, an agent must: Be capable of metacognition, which includes (inter alia): forming dynamic mental goals, i.e., appreciating, adjusting, and revising prior preferences about mental goals. » (2003, p. 501-502, nous soulignons). C'est ce qui nous rend conscient de notre propre existence en tant qu'agent, puisqu'elle nous permet d'agir sur nos propres états mentaux, et qui, par conséquent, rend notre propre identité intuitivement indéniable. Elle nous permet de ressentir une cohérence dans nos états mentaux ainsi que nos actions, ce qui nous permet d'agir dans l'univers social. Proust fournit une définition de l'identité personnelle qui va dans le sens de Parfit, puisque ce qui caractérise l'individu humain non pathologique selon elle, c'est sa capacité à avoir accès à ses états mentaux afin de les réviser de manière à mieux agir dans le monde extérieur. Cette définition permet non seulement de comprendre ce qu'il y a d'intuitif dans le critère psychologique de l'identité personnelle, mais elle permet également de comprendre les mécanismes physiques à l'origine de ce qui nous fait nous considérer en tant qu'individu. Ces mécanismes consisteraient, toujours selon Proust, en deux parties : le niveau du contrôle et le niveau de ce qui est observé. Ces deux niveaux cognitifs correspondraient respectivement au soi idéal et au soi réel – en ce que le contrôle est en vue d'un but et

l'observation consiste en l'évaluation de l'état actuel. Ainsi, ce qui caractériserait l'individu humain, ce serait sa capacité à avoir accès à ses états mentaux, en les observant, et à ensuite pouvoir se fixer des objectifs à atteindre, ce qui pourra être réalisé grâce aux structures de contrôles. Selon Proust, les « soi » résultent de « metacognitive processes in which minds reorganize themselves to be more attuned to their physical and social worlds ». Les processus métacognitifs auxquels Proust réfère sont :

(a) Be capable of metacognition, which includes (inter alia): forming dynamic mental goals, i.e., appreciating, adjusting, and revising prior preferences about mental goals. (b) Form overlapping memories of previous revision episodes. (c) Reorient one's own mental actions on the basis of (a) and (b); revisions are used in the course of planning overlapping future courses of mental actions. (Proust, 2003, p. 502)

En distinguant l'acte mental de l'action, Proust distingue le soi de l'agent et évite ainsi de tomber dans le piège circulaire qui consiste à présupposer l'existence d'un soi pour qu'il y ait agentivité mentale : « Our definition is thus not circular. No self is presupposed before the reflexive¹⁴ intervention of revision gets into play. The self results from this reflexive process, and will stop developing, or decay, if the reflexive process is interrupted, or is reduced. » (Proust, 2003, p. 504) Or, puisque l'identité personnelle à travers le temps (qu'elle soit fondée sur un critère corporel ou psychologique) repose sur l'idée qu'un individu continue d'exister (persiste) à travers le temps, éviter le piège de la circularité soi-agent est primordial. Une manière de récupérer les critères psychologique et corporel en évitant ce piège est de fonder le premier sur l'idée que la continuité de l'individu repose sur sa vie corporelle, donc sur

¹⁴ Il faut comprendre ce terme dans le sens du concept de réflexivité, conformément à la distinction de Seigel (2005).

ses capacités d'agentivité sensorimotrices, et de fonder le deuxième sur l'idée que la continuité de l'individu repose sur sa vie mentale, donc sur ses capacités métacognitives. Dans les deux cas, et tout comme la définition de Proust, le soi (la personne) n'est pas présupposé. Autrement dit, pour éviter de tomber dans ce piège, il faut avant tout se demander « comment sais-je que je suis une personne (que j'ai/je suis un soi)? » Il ne sera pas possible de répondre à la question de la persistance de l'identité personnelle à travers le temps sans présupposer l'existence du soi, et ainsi tomber dans le piège de la circularité soi-agent, sans déterminer avant toute chose qu'est-ce qu'un soi.

What is a person, and how can a person come to know that she is a person? Several answers have been explored by philosophers – having an individual body, and individual brain, having specific introspective access to one's thoughts. They all turned out to be non-starters. A major reason why they do not work, is that they fail to account in a non-circular way for the fact that a person is inherently both a stable and a changing entity; an entity, furthermore, who knows herself as herself. (Proust, 2003, p. 495)

Proust échappe à ce piège en naturalisant la définition de ce qu'est le soi : l'agent ne présuppose pas l'existence d'un soi puisque l'action ne présuppose pas de relation de causalité avec l'agent qui la pose – il peut en effet s'agir d'une action quasi-réflexe en réponse à un stimulus environnemental (ce n'est pas moi qui cligne la paupière quand l'ophtalmologiste envoie un jet d'air dans mon œil, mais elle qui cligne « toute seule »). Selon Proust, pour que l'agent puisse être considéré comme étant une personne, il doit y avoir métacognition permettant une relation réflexive avec soi. Donc, ce qui caractériserait la personne, c'est sa capacité à agir mentalement, c'est-à-dire à accéder à ses états mentaux et de pouvoir les influencer. L'aspect réflexif de l'individu est lié à son identité numérique, en ce que la capacité à réviser ses états mentaux est précisément ce qui garantit son ipséité. À ce sujet, Proust dit :

Note moreover that reflexivity and, consequently, numerical identity are intrinsic to the permanent revisional process in which acting mentally consists. This is crucial to prevent the reduplication problem. Even if, at a given moment, an individual's thought was copied into another's mind, each clone would re-individuate herself through the practical reflexivity that governs her mental actions; as soon as each agent has revised her beliefs to act in her own sphere, with her own body, she has become a person with the double backward/forward dimensions of reidentification that are open to her. (2003, p. 501)

Ainsi, selon cette définition du soi, il doit d'abord y avoir réflexivité pour qu'il y ait un soi; tant qu'il n'y a pas de capacité de révision réflexive, il n'y a pas de soi. C'est pour cette raison que Proust affirme échapper au problème de la circularité. Selon elle, le soi n'est ni une illusion, ni une substance mystérieuse. Il s'agirait plutôt d'un « system of dispositions, socially encouraged and trained, designed to revise beliefs, desires, intentions, and thereby become the actor/goal/target/ of one's own life. » (Proust, 2003, p. 504). Cette version raffinée et naturalisée de la définition du soi évite effectivement le piège de la circularité décrit dans le paragraphe précédent, mais il s'agit d'une définition du soi qui ne décrit, selon nous, que le niveau hiérarchique¹⁵ le plus élevé de celui-ci, à l'image du critère psychologique de l'identité personnelle, ignorant ainsi les niveaux hiérarchiques inférieurs, qui sont tout aussi importants pour l'identité personnelle et qui rendent compte de la réalité intuitive que capture le critère corporel de l'identité personnelle. De plus, le critère d'identité personnelle à travers le temps découlant de la définition du soi dont il est ici question demeure fondé sur une impression de continuité et de connexité, qu'il n'est pas possible de prouver objectivement, d'un point de vue à la troisième personne (c'est-à-dire que nous n'avons pas le moyen de montrer qu'une réelle continuité et connexité correspond à cette

¹⁵ Les niveaux hiérarchiques du soi seront discutés dans la section 2.3.2.

impression). En effet, le soi tel que défini par Proust repose sur la vie mentale et les actions qui en résultent. Le critère de l'identité personnelle que l'on peut en faire découler étant une reformulation du critère psychologique, sachant que sa définition est une version révisée de la théorie du soi de Locke, basée sur la mémoire, n'échappe quant à lui pas au piège de l'intuition dont sont victimes Thomson et Parfit. Malgré cette redéfinition du soi, l'impression d'identité de la personne qui en résulte demeure intuitive – j'ai l'impression d'être la même personne à travers le temps car je ressens une continuité dans mes désirs, mes croyances, mes intentions, etc.

Dans le contexte métaphilosophique naturaliste dans lequel nous plaçons notre enquête sur l'identité personnelle, une faute épistémique est commise lorsque nos postulats philosophiques se basent sur des intuitions. Tel que défendu par Robert Cummins (1998), l'argument du manque de calibration indépendante de l'intuition s'énonce comme suit :

[P1] Nous avons de bonnes raisons de croire le contenu d'une source d'informations donnée seulement si nous possédons une justification indépendante permettant de croire que la source en question est fiable.

[P2] Or nous ne possédons pas de justification indépendante nous permettant de croire que les intuitions sont fiables.

[C] Donc nous n'avons pas de bonnes raisons de croire le contenu de nos intuitions.

L'idée ici n'est pas de déterminer si notre intuition est effectivement ou non une source fiable de connaissance. Il s'agit plutôt de rappeler qu'un postulat tiré d'une vérité intuitive – en ce qu'elle est indéniable phénoménologiquement parlant – est nécessairement épistémiquement circulaire, puisque sa conclusion est contenue dans ses prémisses : « Il m'apparaît indéniable que je suis la même personne donc je suis la même personne (m'apparaît-il indéniablement) ». La seule manière de fonder de manière épistémiquement justifiée une vérité intuitive est d'étudier ses causes (les

mécanismes cognitifs impliqués dans les jugements d'identité personnelle) de manière indépendante à ses effets (notre compréhension intuitive de notre identité personnelle). Par exemple, un objet qui est perçu par l'œil humain comme étant rouge est indéniablement rouge du point de vue phénoménologique. Ceci n'indique pas pour autant que cet objet est « objectivement rouge » (voire qu'il existe une telle chose que le « rouge objectif »). En effet, grâce à l'étude des mécanismes des processus sensoriels, nous savons que ledit objet n'est pas réellement rouge, mais que notre système visuel détermine la couleur phénoménale qu'il assignera à un objet non seulement en fonction des propriétés lumino-réfléctives, etc., de la matière qui le compose, mais également en fonction des conditions lumineuses environnantes (de sorte que deux objets aux propriétés lumino-réfléctives, etc., identiques pourraient être perçus comme étant de couleurs différentes, et inversement deux objets aux propriétés lumino-réfléctives, etc., différentes pourraient être perçus comme étant de la même couleur). Dans un tel cas, le fait de savoir que notre perception du rouge de l'objet est le résultat d'une interaction complexe entre l'état de notre système perceptuel et celui du monde ne change rien au fait que nous le percevons tout de même comme étant rouge. Cela est dû à la transparence phénoménale (Metzinger, 2003). La transparence phénoménale décrit le réalisme de nos expériences phénoménales. Autrement dit, la représentation phénoménologiquement vécue n'est pas vécue en tant que représentation mais bien en tant que réalité. Nos perceptions sensorielles, par exemple, sont phénoménalement transparentes. À l'autre extrémité du spectre se trouve l'opacité phénoménologique, qui elle caractérise la nature représentationnelle de nos expériences mentales. Par exemple, l'expérience phénoménale de nos pensées est opaque, car nous sommes conscients de leur constitution et savons les différencier des expériences qui se passent réellement dans notre environnement. Une bonne façon d'illustrer la différence entre la transparence et l'opacité phénoménale, c'est en pensant à la différence entre des hallucinations résultant d'une expérience psychotique (hallucinations dites « vraies ») versus des pseudo-hallucinations (hallucinoïse, ou hallucination dite « fausse »). Dans le premier cas, ce qui est perçu est tenu pour réel, alors que dans le deuxième cas, ce

qui est perçu est reconnue comme étant factice et résultant d'une illusion. Ainsi, la manière dont on perçoit notre environnement est une vérité intuitive indéniable du point de vue phénoménologique. Or, affirmer que l'objet en question est rouge précisément car nous le percevons comme étant rouge serait une faute épistémique suivant l'argument du manque de calibration indépendante. Cet énoncé serait informationnellement circulaire et il s'agirait là d'une pétition de principe, car sa conclusion est contenue dans ses prémisses – alors on dirait que le fait de le percevoir comme étant rouge est une condition nécessaire au fait que l'objet soit rouge. Si c'est uniquement sur notre perception de l'objet que nous basons notre étude de la nature chromatique de l'objet, il est impossible d'avoir une connaissance fiable et indépendante de celui-ci. Au contraire, il faudrait étudier ses causes indépendamment de ses effets – c'est-à-dire qu'il faudrait investiguer l'objet et ses propriétés indépendamment de la manière dont nous le percevons. Ce n'est qu'une fois que ces deux éléments sont investigués indépendamment qu'ils pourront ensuite être mutuellement utiles à leur étude; le fait que nous voyons l'objet comme étant rouge est utile à la compréhension de la relation entre notre système perceptuel et la nature de l'objet. C'est d'ailleurs précisément ce en quoi consiste la neurophénoménologie. En combinant la neurobiologie et la phénoménologie, la neurophénoménologie tente de résoudre le problème du fossé explicatif (*explanatory gap*), soit la difficulté à expliquer comment nos expériences phénoménales résultent de propriétés physiques. Selon Evan Thompson et al. :

Neurophenomenology asserts that first-person methods are necessary to gather refined first-person data, but not that subjects are infallible about their own mental lives, nor that the experimentalist cannot maintain an attitude of critical neutrality. First-person methods do not confer infallibility upon subjects who use them, but they do enable subjects to thematize important but otherwise tacit aspects of their experience. (2009, p. 71)

Or, nous ne pouvons pas nous contenter de connaître nos intuitions du point de vue de la première personne à propos de l'identité personnelle à travers le temps pour en établir le critère ou pour définir ce qu'est le soi. Il est essentiel, pour des raisons d'intégrité épistémique, d'investiguer la question de manière indépendante et objective avant d'ensuite pouvoir l'ajouter à nos connaissances intuitives.

2.1 Le problème de la persistance de l'identité personnelle à travers le temps reformulé

Jusqu'à maintenant, rien ne nous justifie de considérer le soi en tant qu'entité réelle – nous avons une définition claire de ce qu'est un agent et de comment celui-ci peut en venir à se considérer comme étant une personne, donc comme possédant/étant un soi, mais aucun examen objectif et d'une perspective à la troisième personne ne permet de confirmer le statut ontologique du soi. On peut toutefois considérer le soi comme étant un objet phénoménologique – son existence nous apparaissant phénoménologiquement indéniable. Il est en effet évident que nous ne pouvons pas nous contenter de le considérer comme étant illusoire, puisque phénoménologiquement, il semble être une entité réelle pour toute personne possédant les critères énumérés par Proust et considérés comme en étant les conditions d'existence. Tel que celle-ci le souligne, même lorsqu'un individu croit ou prétend ne pas être lui-même, il est quand même quelqu'un, il affirme tout de même être un soi, avoir une identité personnelle – par exemple, qu'un individu affirme du jour au lendemain être Napoléon ne change pas le fait qu'il se croit assurément être quelqu'un.

Such a reduction does not aim at "doing away" with persons, however. Persons may not be fictions – not only a matter of "self-presentation". For when somebody pretends to be someone he is not, he is still expressing his actual capacity at revising and planning. (2003, p. 504)

Cela étant dit, Proust nous informe uniquement sur les qualités, ou caractéristiques, du soi, mais elle ne nous indique pas les raisons pour lesquelles il faudrait considérer ces qualités comme étant nécessaires et suffisantes pour constituer le soi – si ce n’est parce que nous en avons une idée préconçue, ou phénoménologiquement vécue, et que nous nous contentons d’en analyser les caractéristiques constitutives pour ensuite en faire une description naturalisée. De la même manière, rien ne nous permet de le considérer comme étant une entité ontologiquement réelle. Selon Thomas Metzinger :

Just as current physics is fully compatible with the view that quantum-objects are non-individuals, the corresponding metaphysical underdetermination of cognitive neuroscience certainly makes it tenable that mental, psychological, or phenomenological entities like “the self” are not proper individuals at all: they simply possess no clearly specifiable identity criteria. (2010, p. 280)

Pour sortir de cette impasse épistémique, il faut reformuler la manière dont la question de l’identité personnelle est discutée. Poser la question du critère de persistance de l’identité personnelle en se basant sur nos intuitions, c’est poser la question de manière à ce que nous soyons enclin à adopter une perspective à la première personne ou à la deuxième personne pour y répondre. Pour éviter le piège circulaire de la réponse intuitive, il est essentiel d’également poser la question de manière à faire appel à la perspective à la troisième personne. Les réponses que peuvent nous fournir les perspectives à la première et à la deuxième personne doivent être combinées à celles issues de la perspective à la troisième personne, de manière à se contraindre mutuellement et à nous fournir un portrait plus complet et plus juste, et d’ainsi éviter les observations circulaires. Au lieu de se demander quel est le critère de persistance de l’identité personnelle à travers le temps, il faudrait plutôt et d’abord se demander qu’est-ce que l’identité personnelle et quelles sont les différentes expressions de sa persistance à travers le temps, si persistance il y a?

2.1.1 Distinction entre nécessité phénoménologique et réalité ontologique

L'une des principales caractéristiques du cerveau humain, c'est sa capacité à « boucher les trous » environnementaux sans que l'esprit ne se rende compte de l'illusion (en psychologie, on appelle ce phénomène le « *filling-in* »). Par exemple, nous avons tous un point aveugle, c'est-à-dire qu'une partie de notre rétine ne possède pas les photorécepteurs permettant de percevoir la lumière. Mais, comme l'explique Chris Frith, « We are not aware of this blind spot because the brain makes something up to go into that part of our visual field. Our brain uses the signals from the region immediately around the blind spot to supply the missing information » (2007, p. 135). Il ne s'agit que d'un exemple parmi plusieurs autres. Metzinger affirme à ce sujet que nos cerveaux, plutôt que de simplement percevoir les propriétés des objets, auraient tendance à percevoir les objets comme étant des entités globales, cohérentes et à part entière. Une telle représentation des objets résulterait d'un processus dynamique d'auto-organisation qui lie toutes les propriétés ensemble jusqu'à donner naissance à un objet qui nous semble unifié et être constitué de ces propriétés (2010, p. 282). Le soi tel qu'on le conçoit communément semble être le produit de ce même processus. Or rien n'indique que nous devrions considérer le soi comme étant une entité ou un objet distinct et réel. Au contraire, l'état actuel de nos connaissances en philosophie des neurosciences cognitives¹⁶ nous indique que nous serions justifiés de plutôt le considérer comme étant un simple ensemble de propriétés, le résultat d'un processus par lequel notre cerveau transforme un ensemble de propriétés en objet à part entière n'étant pas garant du statut ontologique de ce dernier. Cependant, tel que précédemment mentionné, il n'en demeure pas moins qu'il existe un soi

¹⁶ Les travaux de Thomas Metzinger (2008, 2010, 2013), entre autres chercheurs, indiquent que le soi est un processus continu et dynamique qui n'a rien de substantiel ni de clairement délimitable et localisable dans le cerveau. Il s'agit plutôt d'un ensemble de propriétés et de processus cérébraux.

phénoménologique, de sorte qu'on ne peut pas simplement affirmer sa non-existence. Une telle chose serait paradoxale, car il est impossible d'échapper à la réalité phénoménale, or nier son existence serait illogique.

Affirmer que le soi n'est pas une entité réelle, mais plutôt un simple ensemble de propriétés, c'est-à-dire qu'il ne constitue pas une catégorie ontologique, ne signifie pas qu'il faille nier l'existence d'un soi en tant qu'objet phénoménologique. Il est d'ailleurs impossible d'imaginer une expérience de pensée au sujet de l'identité personnelle dans laquelle le soi n'existe pas : du moment qu'il est question de générer un scénario, quel qu'il soit, nous manifestons toutes les composantes du soi telles que décrites par Proust, nous empêchant ainsi d'imaginer un réel scénario de non-existence du soi (compris en tant qu'objet phénoménologique). Il serait par conséquent justifié d'affirmer que l'existence du soi est une nécessité phénoménologique. Par contre, comme les données scientifiques l'indiquent, le soi n'est pas pour autant une nécessité métaphysique, bien que nous en ayons l'impression, puisque nous arrivons très bien, jusqu'ici, à étudier les phénomènes cognitifs qu'on lie communément au soi sans avoir à faire appel au concept de soi en tant qu'entité. Il est par conséquent nécessaire, tel que l'indique Metzinger, de distinguer la nécessité phénoménologique de la nécessité métaphysique. Des cas pathologiques extrêmes peuvent nous aider à le faire : par exemple, le patient qui est atteint du syndrome de Cotard aura la ferme conviction d'être déjà mort, alors que, clairement, il ne l'est pas. Dans un tel cas, l'expérience phénoménologique de l'individu en question est réelle, mais ontologiquement parlant, sa croyance est nécessairement fautive, puisqu'il est évidemment vivant. Dans le même ordre d'idées, les cas pathologiques hallucinatoires donnent la vive impression à la personne qui en souffre que toutes ses expériences sensorielles hallucinées sont réelles. Or, bien qu'il s'agisse ici de cas pathologiques, ces cas extrêmes font partie du même spectre que celui dans lequel figure la réalité telle que la personne non pathologique la vit. Ainsi, la réalité phénoménologique vécue n'est pas nécessairement toujours compatible avec la réalité ontologique – et puisque rien ne nous indique jusqu'à maintenant que le soi

soit une entité réelle, il est plus prudent, pour des raisons de simplicité et de parcimonie, de ne pas postuler un réalisme du soi.

Selon Metzinger, notre croyance intuitive voulant que le soi existe nécessairement et qui nous rend logiquement incapable de nier son existence est due à trois facteurs : le facteur de l'architecture cognitive, le facteur de la confusion acte-objet et le facteur évolutif. Le premier facteur fait référence au fait que, comme nous l'avons précédemment mentionné, il n'est pas possible de concevoir le non-soi : « it is a functional feature of our own mental architecture that any attempt to consciously simulate a world in which the no-self hypothesis turns out to be true generates the phenomenology of mental agency, cognitive control, and therefore selfhood. » (2010, p. 293-294) En effet, comme l'a fait remarquer Proust, même lorsqu'on affirme sa propre non-existence, on fait preuve d'agentivité et de contrôle cognitif, soit les caractéristiques même du soi. Quant au second facteur, il s'agit de notre tendance à prendre les processus pour des entités en soi lorsque nous décrivons les caractéristiques de notre soi ou tentons de le décortiquer : « in describing this phenomenology, we are prone to act-object equivocations, because we turn a process into a thing, hypostatizing a phenomenal individual where there is only an intermittent chain of events, using "I" as a designator where it is only an indicator. » (2010, p. 294) Cette critique fait écho à la conception humienne de l'identité personnelle : bien que nous ayons l'impression qu'une entité à part entière et, en quelque sorte, immuable (comme une substance) telle que le soi existe, chaque instant n'est en fait ponctué que d'une série de perceptions, faisant de notre expérience un processus constant de perceptions sensorielles. Le troisième et dernier facteur identifié par Metzinger se rapporte à la manière dont nous ne différencions pas la nécessité phénoménologique des phénomènes vécus de leur nécessité métaphysique :

evolution has shaped our brains in a specific way. We have a strong inbuilt tendency to interpret a functionally grounded phenomenal

necessity as a metaphysical necessity. But of course, the fact that we simply cannot consciously imagine a certain state of affairs has no implications for the deepest structures of reality. Metaphysics is underdetermined by phenomenology. (2010, p. 294)

Autrement dit, le résultat évolutionnaire de notre cognition donne à notre compréhension intuitive du monde une apparence de réalité ontologique. En effet, l'évolution nous a donné une disposition à considérer la délivrance introspective comme la réalité. Toutefois, et tel que précédemment établi, l'information que nous fournit notre intuition phénoménologique est largement insuffisante pour déterminer notre croyance en la nécessité métaphysique, qui se traduit par une impression de réalité ontologique, du soi.

Nous pourrions dire que, puisque l'esprit humain est cognitivement constitué de manière à agir sur ses propres états pour mieux interagir avec le monde extérieur et que c'est ce même monde extérieur qui indique à l'esprit les modifications à apporter à ses états afin de mieux y agir, nous sommes justifiés d'affirmer que l'esprit impose une nécessité phénoménologique qu'on confond avec une réalité ontologique, puisque notre architecture cognitive nous oblige à nous fier à nos perceptions et à les traiter comme si elles étaient toutes réelles – le fait qu'un objet soit réellement rouge ou non importe peu, notre architecture cognitive nous oblige à le percevoir comme étant réellement rouge. Et c'est pour cette raison que c'est tantôt le critère corporel et tantôt le critère psychologique qui est plus intuitif. En effet, il est inconcevable pour l'esprit humain d'imaginer avec une perspective à la première personne sa non-existence ou, à partir d'une perspective à la deuxième personne, la non-existence d'une personne avec qui l'on est en train d'interagir, car c'est grâce à la métacognition, et donc à sa propre personne, que ce scénario est imaginé. Par conséquent, nous nous voyons dans l'obligation d'affirmer qu'il est impossible de nier l'existence des personnes, ou « soi », puisque c'est précisément une personne qui réfléchit à l'existence ou la non-existence des personnes. On pourrait alors supposer que c'est cette disposition intrinsèque à

l'esprit humain qui amène Parfit et Thomson à postuler des critères de l'identité personnelle sur la base d'une pétition de principe, en ce que (1) le critère psychologique ainsi que le critère corporel de l'identité personnelle sont fondés sur l'idée qu'ils sont intuitifs (étant validés par des expériences de pensée qui elles reposent sur nos intuitions) et que (2) il n'est pas possible de postuler quelque chose de contre-intuitif en ce qui a trait à l'identité personnelle, sachant que c'est précisément cette chose intuitive qui chercherait à postuler quelque chose de contre-intuitif. Ce serait pour cette raison que le raisonnement de Thomson et celui de Parfit seraient fallacieux, en ce qu'ils considèrent ontologique une vérité qui n'est au fond que phénoménologique. Toutefois, il faudrait préciser que le fait de postuler ces deux critères comme étant des nécessités phénoménologiques n'est pas problématique en soi, puisqu'elles sont phénoménologiquement vraies. Qui plus est, se limiter à postuler leur justesse qu'en termes phénoménologiques permettrait la possibilité qu'il y en ait plus d'un et éliminerait ainsi le dialogue de sourds qui a lieu entre les partisans de chacun des deux critères. Là où le bât blesse, c'est lorsque cette nécessité phénoménologique est considérée comme étant également une réalité ontologique, ce qui, évidemment, est fallacieux, sachant que ce postulat repose justement sur un raisonnement circulaire, conformément aux trois facteurs énumérés par Metzinger. Or faut-il en conclure que les critères de Parfit et Thomson doivent être rejetés en raison de leurs prémisses circulaires?

Le fait que les critères de l'identité personnelle tels qu'articulés par Parfit et par Thomson soient fondés sur un raisonnement circulaire ne signifie pas pour autant qu'ils ne sont pas valables, en ce qu'ils rendent tout de même compte d'une réalité phénoménologique indéniable. Mais comment deux critères qui s'opposent peuvent tous deux être intuitifs? Pour répondre à cette question, il faudra explorer l'aspect incarné de la cognition. L'esprit est indissociable du corps et vice versa. Sans cette interconnexion, nous n'arriverions pas à interagir avec notre environnement et à planifier nos actions futures : « For our brains, [...] this is not a buzzing blooming

confusion we see around us, but a map of signs about future possibilities. And through this map of future possibilities our bodies are intimately tied to the world immediately around us » (Frith, 2007, p. 98). Ceci implique qu'il suffit de porter attention à un objet dans notre environnement pour que notre cerveau prépare aussitôt notre corps à interagir avec lui. C'est en sens que notre esprit serait intégré à l'environnement et vice versa, faisant de la cognition un ensemble de processus intrinsèquement incarnés. La capacité de préhension humaine illustre bien ce phénomène : « Umberto Castiello and his colleagues ¹⁷ have demonstrated how different objects in a visual scene automatically activate the responses needed to reach and grasp them (the action program) without any conscious intention to act » (Frith, 2007, p. 99). Or nous ne sommes pas conscients de ces mécanismes cérébraux et nous ne nous rendons pas compte que notre connaissance du monde est le fruit d'un apprentissage associatif qui nous donne l'impression qu'il y a nous, en tant que soi, et le monde extérieur, c'est-à-dire le monde physique. Il faut retenir deux choses de ce qui vient d'être établi :

- (1) la mémoire¹⁸ est centrale à notre capacité d'interaction avec le monde physique;
- (2) la distinction entre la vie psychologique et le monde physique, incluant nos propres corps, est dynamique.

En effet, notre interaction avec le monde « extérieur » consiste en une relation bidirectionnelle entre notre cerveau et les informations environnementales, qui seront

¹⁷ Umberto C. (2005). The neuroscience of grasping. *Nature Reviews Neuroscience*, 6(9), 726-736.

¹⁸ Il n'est pas ici question de mémoire explicite, mais plutôt de mémoire implicite, et plus précisément de mémoire procédurale. La mémoire procédurale nous permet de réaliser toutes les tâches quotidiennes qui nous semblent anodines, mais qui résultent en fait d'intégrations complexes. Nous arrivons à les réaliser car, bien que nous n'en soyons pas conscients, nous nous souvenons les avoir déjà réalisées. Ce type de mémoire, et par conséquent ce type de tâches, repose de manière intrinsèque sur l'intégration de l'esprit et du corps - savoir comment faire du vélo est autant une tâche mentale que physique.

encodées, c'est-à-dire retenues en mémoire. Par ailleurs, le simple fait de percevoir fait appel à la mémoire en ce que de les capacités perceptives résultent de la mémoire sensorielle, qui elle consiste en deux autres types de mémoires, soit la mémoire iconique et la mémoire échoïque. Ainsi, la mémoire et nos interactions avec le monde que nous qualifions d'extérieur sont intimement liées, mais cela ne se limite pas aux objets du monde dit physique; la mémoire sémantique est liée à notre vie conceptuelle, c'est-à-dire notre capacité à avoir une conception abstraite du monde sans que ses objets soient physiquement présents. Nous pourrions aller encore plus loin et soutenir une conception étendue de l'esprit, c'est-à-dire étendre l'esprit au-delà de la barrière de la boîte crânienne, et même de la peau. Dans *The Extended Mind*, Andy Clark et David Chalmers illustrent ce cadre de pensée à l'aide de l'exemple suivant : Inga et Otto veulent tous les deux se rendre au musée, qui est situé sur la 53e rue; Inga arrive à s'y rendre car elle avait en mémoire l'adresse; quant à Otto, qui souffre d'Alzheimer, il a pu s'y rendre car il avait inscrit l'adresse dans son carnet, qu'il a toujours avec lui et dans lequel il inscrit toutes les informations dont il doit se souvenir. Clark et Chalmers affirment que dans cette expérience de pensée, Inga ainsi qu'Otto sont allés au musée car ils avaient tous les deux le désir d'y aller ainsi que la croyance qu'il était situé sur la 53e rue. Aussi, de la même manière qu'Inga possédait cette croyance avant même de « consulter » sa mémoire, Otto possédait cette croyance avant même de consulter son carnet. Clark et Chalmers veulent donc faire ici l'analogie entre le rôle que joue la mémoire chez Inga et le rôle que joue le carnet chez Otto : tout comme la mémoire, le carnet contient l'information dont sont constituées des croyances. À l'objection selon laquelle Otto n'a pas de croyances lorsqu'il ne consulte pas son carnet, Clark et Chalmers répondent qu'on pourrait alors affirmer qu'Inga n'a pas de croyances lorsqu'elle n'est pas consciente de ses souvenirs. Et dans le cas où Otto aurait un frère jumeau en tous points identique physiquement, mais qui aurait comme croyance que le musée est situé sur la 51e rue car c'est ce qu'il a inscrit dans son carnet, alors on pourrait très facilement comprendre en quoi les croyances ne sont pas situées dans la boîte crânienne : « The moral is that when it comes to belief, there is nothing sacred about

skull and skin. What makes some information count as a belief is the role it plays, and there is no reason why the relevant role can be played only from inside the body » (Clark & Chalmers, 1998, p. 14). Mais une telle conception de l'esprit est dans une certaine mesure contre-intuitive, car nous avons tendance à concevoir notre esprit, et en particulier notre mémoire, nos désirs et nos croyances, comme étant uniques et comme étant plus fiables que les autres ressources, car en sûreté – contrairement au carnet, qui ne serait qu'un support « externe » et qu'on pourrait perdre ou qui pourrait être détruit. À cette objection, Clark et Chalmers répondent que « these various small differences between Otto's and Inga's cases are all shallow differences. To focus on them would be to miss the way in which for Otto, notebook entries play just the sort of role that beliefs play in guiding most people's lives » (Clark & Chalmers, 1998, p. 16).

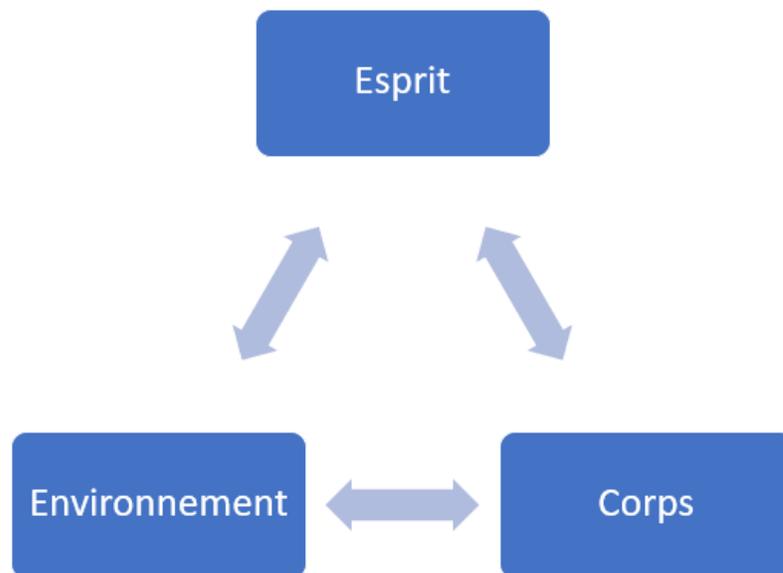


Figure 2.1 L'esprit étendu

Or selon une telle conception, la querelle corps-esprit comme critère de l'identité personnelle perd un peu de son tranchant. Sans éliminer les différences entre les caractéristiques proprement psychologiques et le rôle du corps dans la continuité de l'identité, il n'en demeure pas moins que cette conception brouille quelque peu les démarcations des deux critères, en ce qu'ils sont désormais interdépendants. Bien que la perspective incarnée n'élimine pas l'importance de la continuité des souvenirs, désirs, croyances et autres caractéristiques psychologique, elle met néanmoins en lumière la non-opposition de celui-ci au critère corporel. Au contraire, et dans le même ordre d'idée que ce qui a été affirmé dans les paragraphes précédents, ces deux critères sont complémentaires. Bien que la manière dont nous concevons phénoménologiquement la réalité nous donne l'impression d'un dualisme corps-esprit, rien n'empêche qu'une conception incarnée et étendue de la cognition et de l'esprit soit ontologiquement possible. Il suffit de considérer le rôle que jouent les téléphones intelligents dans nos vies en tant que ressources sur lesquelles nous nous fions et sur lesquelles reposent plusieurs de nos désirs, de nos croyances, mais également de nos souvenirs.

2.1.2 Le soi hiérarchiquement construit : récupération des deux critères intuitifs

Pour mieux comprendre l'interaction entre le soi et l'environnement extérieur, et pour une meilleure compréhension de la nature du soi lui-même qui nous sera utile tout au long du mémoire, il est pertinent d'en présenter une analyse détaillée, exposant sa nature hiérarchique et processuelle. Il est essentiel de tenir compte des divers niveaux du soi, à commencer par les niveaux hiérarchiques inférieurs, car cela permet de comprendre ce qu'il y a d'intuitif à chacun des critères de persistance de l'identité personnelle à travers le temps, en ce que chacun capture la réalité de l'un des niveaux hiérarchiques du soi vécu phénoménologiquement.

Le caractère intuitif du critère corporel de la persistance de l'identité personnelle à travers le temps peut s'expliquer par la manière dont nous avons conscience de notre propre corps. Tel que précédemment mentionné, l'agentivité sensorimotrice est

centrale à la continuité corporelle. On peut être conscient de soi-même de manière réflexive, c'est-à-dire, entre autres, de manière volontaire, comme c'est par exemple le cas lorsqu'on nous réfléchissons à nos intentions ou à nos actions. Mais on peut également être conscient de soi-même de manière pré-réflexive, en ce que le sujet (non pathologique) considère, ou plutôt « sait », être le sujet de son expérience. Autrement dit, ce type de conscience est simplement caractérisé par le sentiment d'être celui à qui ses expériences appartiennent (*mienneté*¹⁹). La conscience pré-réflexive du corps se divise en deux formes distinctes : le corps transparent et le corps performatif. Selon Dorothée Legrand :

The body is transparent in the sense that one looks through it to the world. At this level, pre-reflective bodily experience is precisely the experience of the world as given through the "transparent body". The latter is not perceived as an object but experienced specifically as a subject perceiving and acting, that is, as in-the-world. I take this claim to be a phenomenal "evidence" in the simple sense that one can hardly argue for or against it. (2007, p. 504)

Autrement dit, le corps transparent est le médium par lequel le monde est vécu. Quant au corps performatif, il n'est ni objet intentionnel de l'attention, ni le médium transparent de l'expérience. Il s'agit plutôt du corps vécu lors la performance technique (quotidienne ou experte); pendant l'activité de performance, on est conscient de la présence et de l'activité du corps, sans pour autant qu'il ne faille y porter attention (sans que le corps soit « opaque »). C'est le cas lors des performances de danse professionnelle, dans lesquelles le danseur n'a pas besoin d'être attentif aux membres de son corps afin que ceux-ci réalisent les mouvements désirés, ni même à son corps

¹⁹ Le concept de *mienneté* est ici utilisé au sens de *ownership*, tel qu'utilisé par Limanowski & Blankenburg (2013)

en général, mais il demeure en tout temps conscient de sa présence, puisqu'il le contrôle. À ce sujet, Giovanna Colombetti affirme que « The performative body is mainly constituted by prereflective proprioceptive and kinesthetic sensations, namely, sensations of bodily position and movement that are not attended. » (2007, p. 118). Ainsi, la différence entre le corps transparent et le corps performatif réside dans le fait que le premier est présent, bien que de manière non intentionnelle, alors que le dernier est plutôt absent, quasi invisible.

La conscience réflexive de mon propre corps correspond au corps comme objet intentionnel d'attention, et en être consciente nécessite une conscience observationnelle de celui-ci. Au contraire, tel que Shaun Gallagher l'explique :

The sense of ownership... does not require an explicit or observational consciousness of the body, an ideational, third-person stance in which I take my body as an object. Rather it may depend on a non-observational access that I have to my actions, an access that is most commonly associated with a first-person relationship to myself. In non-observational self-awareness I do not require the mediation of a perception or a judgment to recognize myself as myself. I do not need to reflectively ascertain that my body is mine, or that it is my body that is in pain or that is experiencing pleasure. (2005, p. 29).

Autrement dit, le sentiment de *mienneté* ne nécessite pas la médiation d'une quelconque perception ou jugement : à l'image de Thomson qui affirmait que l'on est son corps, le sentiment de *mienneté* corporelle tel que décrit ici par Gallagher est intrinsèque et direct. Selon Colombetti, du moment que notre corps est l'objet intentionnel de notre attention, notre conscience de notre propre corps est réflexive. Cela est aussi inversement vrai : pour être conscient de son propre corps de manière réflexive, c'est-à-dire de manière volontaire et consciente, il faut prendre le corps comme objet intentionnel de l'expérience. Cette relation à double sens ne s'applique toutefois pas au rapport qui existe entre le corps comme étant ce par quoi autre chose

est ressentie et la conscience pré-réflexive de soi. En effet, du moment que le corps agit comme intermédiaire pour le sentiment (*feeling*) d'autre chose, nous avons conscience de manière pré-réflexive de notre corps, en ce que la sensation en question est tacitement vécue comme étant celle du corps qui la vit.

La conscience pré-réflexive de soi est, selon Legrand, au fondement de la conscience réflexive, c'est-à-dire que cette dernière ne peut avoir lieu sans qu'il n'y ait d'abord conscience pré-réflexive. Il s'agirait, d'un point de vue hiérarchique du soi, du niveau inférieur et fondamental des niveaux supérieurs, réfléchitifs du soi. Dans le même ordre d'idée, Sartre a affirmé au sujet de la conscience pré-réflexive ce qui suit : « Cette conscience (de) soi, nous ne devons pas la considérer comme une nouvelle conscience, mais comme le seul mode d'existence qui soit possible pour une conscience de quelque chose. » (1943, p. 20)

On peut donc comprendre en quoi le critère corporel tel que décrit par Thomson est intuitif : nous sommes toujours incarnés et c'est précisément le caractère incarné de notre existence qui fait que nous savons en tout temps que nous sommes nous-même. Le fait qu'il n'y ait pas de coupure temporelle perçue entre chaque instant où nous vivons notre propre corps nous permet de nous considérer comme étant les mêmes à travers le temps, sur la base de notre continuité corporelle. Une bonne manière d'illustrer cette auto-concordance avec soi en tant qu'individu incarné est en citant les mots de Gareth Evans : « I do not move myself, I myself move » (1982, p. 207).

Ainsi, nous pouvons voir, en considérant la nature hiérarchique du soi, que les deux critères intuitifs de l'identité personnelle, soit les critères psychologique et corporel, sont en fait complémentaires et ne peuvent être considérés indépendamment l'un de l'autre dans l'investigation de la persistance de l'identité personnelle à travers le temps. Il faudrait probablement parler *des* critères de persistance de l'identité personnelle et non plus *du* critère de persistance de l'identité personnelle. Grâce à la version révisée

de la définition lockéenne du soi de Proust, nous savons en quoi la métacognition, et par conséquent, la vie mentale, est une composante essentielle de l'identité personnelle à travers le temps. Nous savons toutefois également, grâce aux thèses énoncées ainsi qu'aux travaux sur le soi minimal pré-réflexif, à quel point le corps est un élément fondamental de notre identité. Alors que le premier est garant de notre même à travers le temps (je me souviens de qui j'étais et je me projette dans l'avenir), le second est garant de notre ipséité (je *suis* en tout temps mon corps, peu importe mon niveau attentionnel).

CHAPITRE III

L'IDENTITÉ PERSONNELLE CONÇUE COMME LE RÉSULTAT DE LA MODÉLISATION HIÉRARCHIQUE DE SOI DANS LE MONDE

La complémentarité des deux critères de l'identité personnelle, soit le critère psychologique et le critère corporel, sera maintenant expliquée en termes de modélisation du sentiment de *mienneté*²⁰, car c'est de ce point de départ hiérarchique que l'identité personnelle procède. En effet, tel que discuté dans le chapitre précédent, le sentiment de *mienneté*, via la pré-réflexivité corporelle, est à la base de toute la complexité de la conscience de soi - il ne peut y avoir d'identité personnelle sans qu'il n'y ait d'abord pré-réflexivité, c'est-à-dire sentiment de *mienneté*. Celui-ci est une caractéristique essentielle de l'expérience phénoménale et il accompagne par conséquent tous les niveaux du soi, du plus simple au plus complexe. C'est précisément de cette complexité qu'il sera question dans ce chapitre, soit des divers niveaux hiérarchiques du soi, allant du niveau minimal correspondant davantage à la corporalité, point d'ancrage du sentiment de *mienneté*, au niveau supérieur faisant plutôt référence à l'aspect psychologique (cognitif) du soi. Cette hiérarchie est graduelle et chacun de ses éléments dépend de celui qui le précède.

²⁰ Tel que nous l'avons défini dans le deuxième chapitre de ce mémoire, le sentiment de *mienneté* est ici utilisé au sens de « *ownership* » (Limanowski & Blankenburg, 2013)

Mais comment en arrive-t-on à être conscient de soi? À se considérer comme étant (ou ayant) un soi? Cette question est fondamentale, car il ne peut y avoir identité personnelle sans qu'il y ait d'abord conscience de soi. Dans le cadre théorique du traitement prédictif, on explique la conscience de soi en termes de modèles internes²¹, permettant de se distinguer d'autrui et, par conséquent, d'interagir avec le monde extérieur. Ce processus a lieu par voie d'inférence et d'autoreprésentation. Nous verrons dans ce chapitre le rôle du corps et de la cognition dans la hiérarchie du soi, pour ensuite discuter du caractère probabiliste et inférentiel de celle-ci. Nous verrons de quelle manière la distinction soi-autrui s'opère en début de vie et comment ce processus d'auto-identification peut facilement être altéré lorsque certaines conditions sont réunies. Pour illustrer cette possibilité d'altération, nous présenterons deux expériences d'illusions corporelles démontrant le caractère parfois illusoire de l'identification de soi.

3.1 La hiérarchie du soi

Tel que précédemment affirmé, suivant Metzinger 2008, 2013b, le soi, ou plutôt le sentiment (intuition) d'être/posséder un soi, est le résultat d'un processus hiérarchique. Pour le dire simplement, *le soi est hiérarchique*. Il se subdivise en divers niveaux, dont chacun résulte et dépend de celui qui le précède. Le niveau élémentaire et fondamental de cette pyramide hiérarchique du soi, celui dont tous les autres niveaux procèdent, est celui du *soi minimal phénoménal*.

²¹ Il est important de noter que la version de la théorie du traitement prédictif présentée ici fait partie des interprétations internalistes (Hohwy, 2013), que l'on oppose généralement aux interprétations externalistes de la théorie (Clark, 2016).

3.1.1 Au-delà de la dichotomie corps-esprit

Le sentiment minimal de soi (*minimal selfhood*), c'est-à-dire la représentation de soi, ou autoreprésentation, au niveau de base corporel et dans l'immédiat, dépend selon Metzinger d'un élément très simple et élémentaire : l'autolocalisation spatio-temporelle transparente (Metzinger, 2013b). Ce type d'autoreprésentation se distingue des types de niveaux hiérarchiquement supérieurs, en ce qu'ils sont corporellement ancrés et se situent dans l'« ici et maintenant », par opposition aux types d'autoreprésentation hiérarchiquement supérieurs, qui, eux, sont plutôt d'ordre cognitif et reposent davantage sur des projections temporellement étendues. Il s'agit selon lui d'une condition nécessaire et suffisante de l'auto-identification consciemment et phénoménologiquement vécue. Comme nous l'avons vu dans le premier chapitre de ce mémoire (voir note de bas de page 13), la transparence phénoménale caractérise le réalisme de nos expériences phénoménales. Celle-ci fait référence à l'état dans lequel la représentation phénoménologiquement vécue ne l'est pas en tant que représentation mais bien en tant que réalité. Metzinger définit la transparence phénoménale comme suit :

“Transparency” is a property of conscious representations, namely, that they are not experienced as representations. Therefore, the subject of experience feels as if being in direct and immediate contact with their content. Transparent conscious representations create the phenomenology of naïve realism. An opaque phenomenal representation is one that is experienced as a representation, for example in pseudo-hallucinations or lucid dreams. There exists a graded spectrum between transparency and opacity, determining the variable phenomenology of “mind-independence” or “realness.” Unconscious representations are neither transparent nor opaque. (Metzinger, 2013b, p. 2)

La notion d'« indépendance de l'esprit » (*mind-independance*) à laquelle réfère Metzinger pour décrire la transparence phénoménale se définit comme caractérisant tous les phénomènes existant en dehors de notre esprit, c'est-à-dire des phénomènes

qui n'ont pas été *créés* par notre esprit. Cette notion s'oppose à celle de « dépendance de l'esprit », qui elle décrit les phénomènes créés par notre esprit, tels que les pensées, les rêves, etc., et qui sont vécus comme *étant* des représentations. Tout ce qui est phénoménalement transparent est donc vécu comme étant indépendant de l'esprit et tout ce qui est phénoménalement opaque est dépendant de l'esprit – entre ces deux extrêmes se trouve toutefois tout un spectre graduel d'états phénoménaux. Par exemple, comme nous l'avions illustré précédemment, les hallucinations résultant d'un état psychotique sont phénoménalement transparentes, alors que les fausses hallucinations, telles que celles que nous vivons au quotidien lorsque nous croyons avoir aperçu ou entendu quelque chose, sont phénoménalement opaques, c'est-à-dire que nous sommes conscients qu'il s'agit d'une illusion.

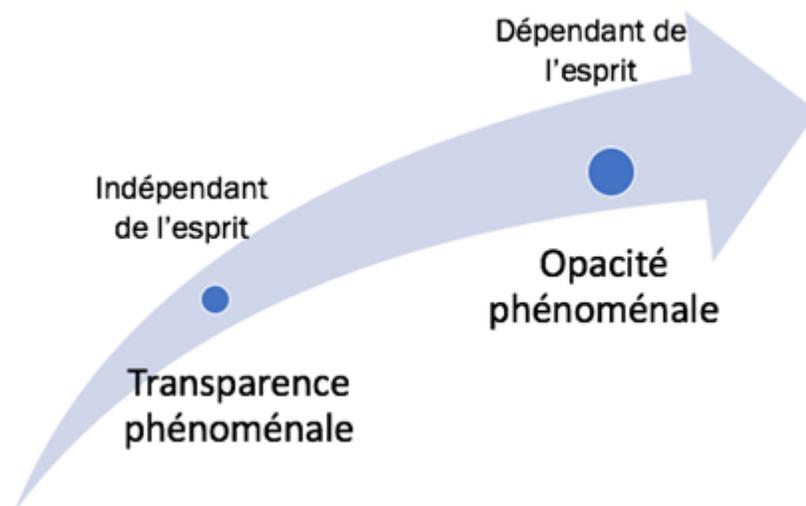


Figure 3.1 Le spectre du sentiment de réalité, de la transparence à l'opacité phénoménales

Ainsi, le soi phénoménal, c'est-à-dire la forme la plus élémentaire de perspective à la première personne, émerge aussitôt qu'il y a capacité à s'autolocaliser spatio-temporellement et de manière transparente, c'est-à-dire à s'auto-représenter consciemment sans que cela ne soit vécu comme étant une représentation. Par autolocalisation spatio-temporelle, nous entendons ici la capacité à localiser son propre corps dans l'espace, mais de manière immédiate et non-conceptuelle. Cette capacité n'implique pas la capacité à s'autolocaliser de manière conceptuelle ni de posséder un niveau de soi hiérarchiquement supérieur. C'est ce que Metzinger appelle le soi minimal phénoménal (*minimal phenomenal self*). Il s'agit du point d'ancrage des formes plus complexes du sentiment de soi. Son existence est nécessaire à l'émergence du sentiment de subjectivité et à la réflexivité.

Comme nous l'avons vu dans le premier chapitre, 1) le soi est phénoménologiquement indéniable, 2) il est hiérarchique – son niveau minimal repose sur une pré-réflexivité corporelle et 3) son niveau supérieur est cognitif. Metzinger et Blanke (2008) décomposent la constitution du soi en trois formes de perspectives à la première personne : la perspective à la première personne faible (*weak first-person perspective* ou *weak IPP*), la perspective à la première personne forte (*strong first-person perspective* ou *strong IPP*) et la perspective à la première personne cognitive (*cognitive first-person perspective* ou *cognitive IPP*). La perspective à la première personne faible fait référence au point d'origine perceptuel. Il s'agit de la perspective géométrique à partir de laquelle l'individu perçoit et possède une représentation globale incarnée de lui-même, lui permettant d'interagir avec son environnement. Elle est constitutive du soi minimal phénoménal et est, par conséquent, nécessaire à la

constitution des niveaux plus élevés de soi, mais n'est pas en elle-même subjective²². Quant à la perspective à la première personne forte, il s'agit de la perspective à la première personne faible additionnée d'un sentiment minimal de soi vécu phénoménologiquement (via le soi minimal phénoménal) – il résulte de la capacité à s'auto-représenter comme étant (attentionnellement) dirigé vers un objet (externe ou son propre corps vécu dans son ensemble). Finalement, la perspective à la première personne cognitive correspond au niveau hiérarchiquement supérieur du soi phénoménal, qui se caractérise par le fait de posséder un concept de la perspective à la première personne forte appliqué à soi. Il se caractérise par la capacité à référer à soi-même en tant que sujet de l'expérience, en utilisant l'indexical « je ». Il s'agit donc de posséder un concept de soi que l'on peut appliquer à soi-même, ce qui implique une forme de capacité réflexive.

A strong 1PP occurs when the model of the organism as a whole, given through MPS, is represented as being directed at an object component (including, potentially, the body itself). A cognitive 1PP appears when a system possesses an abstract concept of itself as a subject of experience, and is able to apply this concept to itself. (Metzinger, 2013b, p. 3)

²² Pour qu'il y ait subjectivité, il doit avant tout y avoir *soi minimal phénoménal*. Or la perspective à la première personne faible est constitutive du soi minimal phénoménal. Il n'y a donc pas encore de subjectivité au stade hiérarchique de la perspective à la première personne faible.

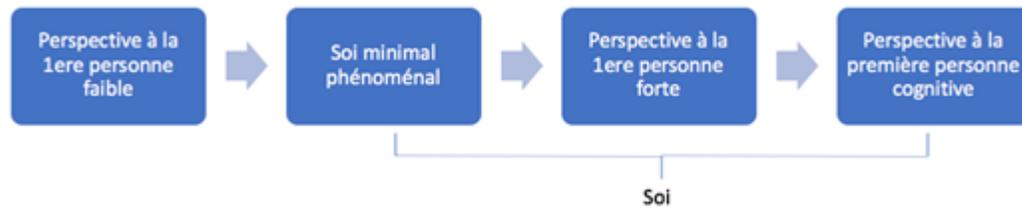


Figure 3.2 Division hiérarchique des niveaux du soi

La base du soi (expérience consciente de soi ou conscience de soi) commence donc au moment où il y a expérience phénoménale. Cet état précède l'expérience subjective et la capacité à agir intentionnellement dans le monde. Il se caractérise par le contenu phénoménal vécu qualitativement et dans le moment présent, et n'implique aucune récursivité ni aucune réflexivité. Il s'agit du soi *dans* le monde, *ici et maintenant*. Il s'agit du monde et de soi vécus comme *étant* la réalité, et non pas comme étant une *représentation* de la réalité. Or, tel que précédemment établi, il est impossible de rendre compte de ces états de manière fidèle. Baser un critère de l'identité personnelle à travers le temps sur nos observations subjectives de notre propre identité personnelle (sur la base d'expériences de pensée auxquelles on répond à partir de notre perspective à la première ou à la deuxième personne) ne peut être épistémiquement valable puisque notre conclusion (critère d'identité) serait contenue dans notre prémisse (notre propre impression d'identité à travers le temps), tel que mentionné précédemment. Cette impossibilité de rendre fidèlement compte de son expérience phénoménale est ce qu'on appelle le *problème du rapport verbal* [*the problem of report*]²³, soit le fait que la seule manière d'avoir accès aux expériences à la première personne d'autrui, c'est via le

²³ Traduction basée sur celle de Stanislas Dehaene (2018) du terme « *report* ». Thomas Nagel (1974, 1986) argue que l'expérience subjective, de par sa nature, ne peut être considérée de manière objective. Selon lui, l'expérience subjective ne peut être réduite à une explication matérialiste et est inaccessible à tous sauf la personne à qui elle appartient.

rapport verbal de celles-ci par la personne à qui elles appartiennent, ce qui engendre une perte qualitative et informationnelle considérable – l’expérience phénoménale ou subjective semblant infiniment plus riche que ce que nous sommes capables de rapporter verbalement, en plus de n’exister que dans le moment présent (elle est automatiquement altérée par l’interprétation et la remémoration une fois le moment passé). Dans le même ordre d’idées, William James affirmait dans *Principles of Psychology* (1918) que rien ne nous permet d’affirmer que nous vivons plus d’une fois la même sensation corporelle et c’est précisément pour cette raison qu’aucune valeur épistémique ne peut être accordée au rapport spontané de ceux-ci : « there is no proof that the same bodily sensation is ever got by us twice. What is got twice is the same object [...] This is what makes off-hand testimony about the subjective identity of different sensations well-nigh worthless as a proof of the fact. » (p. 231)

Tel que précédemment établi, il doit nécessairement y avoir un point de référence égocentrique permettant l’auto-identification pour qu’il y ait soi minimal phénoménal. Dans le cas des êtres humains, ce point de référence est le corps, mais rien n’empêche que ce point puisse être la carcasse d’une machine (pensons à Robocop) ou une quelconque structure physique agissant comme point d’origine d’une machine intelligente advenant le cas où nous réussissons à créer une intelligence artificielle suffisamment avancée. Il faut en effet comprendre que, pour Metzinger, la capacité à s’autolocaliser spatio-temporellement à elle-seule suffit à donner lieu à un sentiment de soi minimal phénoménal. Or cette condition peut être hypothétiquement et potentiellement remplie par n’importe quel type de cadre de référence égocentrique. Quoi qu’il en soit, en ce qui a trait aux êtres vivants possédant un soi minimal, ce cadre de référence égocentrique, correspondant à la perspective à la première personne faible, est le corps. Celui-ci est donc le point d’origine de toutes les formes de sentiment de soi (*selfhood*), de la plus minimale à la plus complexe, soit du soi minimal phénoménal, qui est ancré dans l’expérience corporelle pré-réflexive, à la perspective à la première personne cognitive, qui se caractérise par une capacité métacognitive et réflexive.

MPS is the central enabling condition for having a subjective, consciously experienced 1PP on the weak reading; it is a necessary (but not sufficient) condition on two stronger readings of the term “1PP”. Having MPS is a necessary condition for the strong (i.e., attentional and/or cognitive) 1PP, but not for the weak 1PP, because the weak 1PP is itself a constituting factor for MPS. The notion of “embodiment” is of central relevance for MPS, as is the question of how critical global bodily properties like self-identification, self-location and 1PP can be grounded and functionally anchored in the brain (Metzinger, 2013b, p. 3)

La dichotomie corps-esprit n'est, par conséquent, plus pertinente lorsque considérée comme opposition théorique en ce qui a trait à la persistance de l'identité personnelle à travers le temps. La manière dont le critère corporel de l'identité personnelle a été défendu par Thomson se focalise sur la nature « ici et maintenant » du sentiment de soi. Celle-ci affirmait ce qui suit : « I feel inclined to think that this fleshy object (my body is what I refer to) isn't something I merely currently inhabit: I feel inclined to think that it is me [...] Certainly we all, at least at times, feel inclined to think that we are not merely embodied, but that we just, all simply, are our bodies. » (Thomson, 2008, p. 155) Cette description correspondrait au soi minimal phénoménal, en ce qu'elle décrit le caractère phénoménologiquement transparent de l'expérience de soi en tant qu'individu incarné ainsi que la dimension d'auto-identification essentielle au sentiment de soi. Étant donné la nature hiérarchique du soi, le critère corporel de l'identité personnelle ne peut être écarté, ni peut-il être opposé à un critère cognitif, qui est hiérarchiquement plus élevé, mais qui est nécessairement lié au critère corporel, sachant qu'ils font tous deux partie d'une même hiérarchisation de niveaux de complexité du sentiment de soi. Il en va de même pour le critère psychologique de la persistance de l'identité personnelle à travers le temps, qu'il s'agisse de la version réductionniste défendue par Parfit ou de la version naturalisée défendue par Proust. Les deux articulations de ce critère rendent compte d'un niveau plus complexe et hiérarchiquement supérieur du sentiment de soi. Ces deux redéfinitions du critère lockéen de l'identité personnelle sont basées sur la notion d'agentivité cognitive, en ce qu'elles reposent sur l'idée que nous sommes les mêmes à travers le temps en fonction

de la continuité entre nos états cognitifs (ou psychologiques) passés et futurs. Cela dépend de nos capacités métacognitives, qui nous permettent de moduler nos états futurs en fonction de nos états passés. Le tableau ci-dessous, tiré de l'article *Why are dreams interesting for philosopher? The example of minimal phenomenal selfhood, plus an agenda for future research* (Metzinger, 2013b), liste les composantes de la conscience de soi. Il détaille des éléments de la conscience de soi qui ne sont pas abordés dans ce mémoire en raison de sa taille limitée. Les éléments en question sont l'agentivité corporelle, l'agentivité attentionnelle, l'agentivité cognitive et le modèle de l'agent épistémique. Les trois premiers se caractérisent par la capacité à s'auto-identifier en tant qu'agent possédant une image corporelle : l'agentivité corporelle correspond au contrôle de soi causal, l'agentivité attentionnelle au contrôle de soi attentionnel et l'agentivité cognitive au contrôle de soi cognitif de haut niveau. Quant au modèle du soi épistémique, il se caractérise par la capacité à s'auto-représenter en tant qu'entité individuelle capable d'avoir un contrôle sur elle-même.

1. **MPS** (*minimal phenomenal selfhood*; Blanke and Metzinger, 2009)
Paradigmatic autophenomenological report: <**I am this, here and now.**>
 Self-identification as *transparent spatiotemporal self-location*:
 - a. A conscious self-representation that is not experienced as a representation.
 - b. Satisfaction of transparency-constraint leads to the phenomenology of identification.
 - c. Data from bodiless dreams and asomatic out-of-body experiences show that pure spatiotemporal self-location suffices for a robust sense of self (sections Asomatic OBEs and Bodiless dreams).
 - d. MPS can be experimentally controlled (e.g., De Ridder et al., 2007; Ehrsson, 2007; Lenggenhager et al., 2007)
 - e. Self-location in an egocentric frame of reference creates a weak, purely geometrical first-person perspective (1PP; see below), which is a necessary constituent of MPS.
 - f. Once MPS has been established, it can function as the origin of richer and more complex forms of 1PP
2. **BA** (*bodily agency*)
Paradigmatic autophenomenological report: <**I am an embodied agent.**>
 Self-identification with an extended body image, plus causal self-control.
 - a. A spatially extended, conscious body-representation that is not experienced as a representation.
 - b. A holistic representation of the system as an entity that acts physically, for example by changing its position in space; and/or
 - c. has the *potential* for physical agency.
3. **AA** (*attentional agency*)
Paradigmatic autophenomenological report: <**I am a self in the act of controlling its attentional focus.**>
 Typically, self-identification with an extended body image, plus attentional self-control.
 - a. A conscious self-representation as an entity currently controlling mental resource allocation; and/or
 - b. having the *potential* for controlling subsymbolic mental resource allocation.
 - c. Data from bodiless dreams and asomatic out-of-body experiences show that pure spatiotemporal self-location can coexist with AA (sections Asomatic OBEs and Bodiless Dreams).
4. **CA** (*cognitive agency*)
Paradigmatic autophenomenological report: <**I am a thinking self.**>
 Typically, self-identification with an extended body image, plus high-level cognitive self-control.
 - a. A conscious self-representation as an entity currently controlling high-level cognitive processing, including quasi-symbolic, conceptual, and propositional contents; and/or
 - b. having the *potential* for controlling high-level cognitive processing, including quasi-symbolic, conceptual, and propositional content.
 - c. Data from bodiless dreams and asomatic out-of-body experiences show that pure spatiotemporal self-location can coexist with CA (sections Asomatic OBEs and Bodiless Dreams).
5. **EAM** (*epistemic agent model*; see section on *Bodiless Dreams*)
Paradigmatic autophenomenological report: <**I am a knowing self.**>
 Conscious self-representation of the system as an individual entity capable of epistemic self-control:
 - a. as currently standing in and/or actively constructing knowledge relations to certain parts of the world; and/or
 - b. as having the *ability* to actively establish such relations.
 - c. AA is sufficient for having an EAM, CA is not a necessary condition.
6. **1PP** (*first-person perspective*; Metzinger, 2003a,b, 2006, 2007; Blanke and Metzinger, 2009)
Paradigmatic autophenomenological report: <**I am directed at the world.**>
 Conscious representation of the system as currently being *directed* at the world, either by bodily or by epistemic agency. Comes in many variations, paradigmatic examples are:
 - a. Being directed at a physical goal-state via ongoing bodily agency (BA).
 - b. Being directed at a perceptual object via high-level attentional control (AA).
 - c. Being directed at abstract cognitive content (e.g., concepts or propositions) via cognitive self-control, quasi-symbolic thought, etc. (CA).

Figure 3.3 Tirée de Metzinger, 2013b – énumération de composantes hiérarchiques du soi

Ainsi, il est important de prendre en considération le niveau hiérarchique du soi qui nous intéresse lorsque nous évaluons le critère de persistance de l'identité personnelle. Il peut s'agir d'un soi minimal phénoménal (pré-réflexif), auquel cas c'est la corporalité qui est le critère de persistance d'intérêt. Il peut également s'agir d'un soi

réflectif de plus haut niveau, un agent épistémique capable de se concevoir comme un sujet et de posséder un concept de lui-même, auquel cas l'esprit est ce qui est d'intérêt.

3.1.2 Modélisation de soi et du monde

Tel que précédemment mentionné, nous construisons des modèles de nous-même ainsi que du monde externe afin de pouvoir interagir avec notre environnement. Ce processus s'opère par une relation bilatérale entre soi et le monde externe, de sorte que des *inputs* sensoriels en provenance de ce qui nous entoure ainsi que de notre propre corps sont traités par notre cerveau, qui lui crée à partir de ces deux sources un modèle du monde permettant de le rendre intelligible. Pour le dire simplement, le cerveau, afin de déterminer la cause de l'information sensorielle qu'il perçoit, crée une représentation de ladite cause. Ce processus est toutefois actif, en ce que la représentation en question peut en tout temps être révisée, à la lumière d'une nouvelle information révélant une nouvelle cause. La somme des nouvelles informations permet de trouver la cause probable des données perçues, donnant ainsi lieu à des représentations actualisées. Dans la conception bayésienne de ce processus d'actualisation, cette opération se réalise grâce à la jonction de deux éléments essentiels au processus de création de modèles hiérarchiques de soi et du monde, soit le *prior* et le *likelihood*, c'est-à-dire la croyance initiale et la probabilité que celle-ci concorde avec la nouvelle information obtenue. C'est donc par un processus d'actualisation de la croyance initiale (*prior*), à la lumière de la probabilité de concordance de celle-ci avec la nouvelle information obtenue, qu'il y a inférence. Le résultat de l'opération est appelé *posterior*, c'est-à-dire, dans le cas de la construction d'un modèle du monde, ce qui est maintenant considéré comme étant la cause de l'*input* sensoriel (Clark, 2013; Hohwy, 2013; Seth, 2013; Wiese & Metzinger, 2017) :

Formally, this update involves computing an a posteriori distribution (which is also just called the posterior). The posterior is obtained by combining an a priori distribution (also just called the prior) with a

likelihood. The prior codes the information the agent already has; the likelihood codes how the domain about which the agent already has information is related to the domain of the new information obtained. A nice feature of Bayesian inference is that it can reduce uncertainty. Formally, this means that the posterior often has a lower variance – is more precise – than the prior. (Wiese & Metzinger, 2017, p. 6)

Dans la conception hiérarchique des modélisations bayésiennes du monde, les modèles qui résultent de cette opération sont progressivement organisés sur différentes échelles temporelles : à la base de cette hiérarchie, on retrouve les représentations opérant rapidement (telles que les modalités sensorielles) et aux niveaux supérieurs on retrouve les représentations qui opèrent à une échelle temporelle plus longue (telles que les représentations conceptuelles) (Hohwy, 2017). Et entre les deux, il y aurait des représentations ni totalement sensorielles, ni totalement conceptuelles, pensons par exemple aux représentations d'affordances, ressenties de manière pré-réflexive. Notre modélisation de soi et du monde suit donc ce même principe : l'individu possède un modèle interne hiérarchique de son propre corps, ce qui lui permet de minimiser les erreurs de prédiction. À l'échelle temporelle courte, nous nous représentons les causes de nos perceptions sensorielles (niveau hiérarchique inférieur et corporel), et à l'échelle temporelle longue, nous inférons nos états futurs (niveau hiérarchique supérieur et cognitif), et ce processus est itératif, en ce que le niveau hiérarchique supérieur cognitif contribue à la minimisation des erreurs de prédiction au niveau hiérarchique inférieur et corporel :

As the human self-model integrates different predictive horizons, different timescales generate different types of nested self-representational content: the body-model predicts somatosensory events within very small time-windows of only a few hundred milliseconds, whereas the autobiographical self-model predicts events which may lie years ahead, including the organism's eventual death. The cognitive layer not only biases affordance competition at lower hierarchical levels, it crucially allows an agent to adaptively destroy and create new

cognitive affordances in order to realize long-term goal-representations.
(Metzinger, 2017b, p. 26)

Selon Karl Friston (2017), l'être humain en vient à construire son modèle du soi, dès les premières semaines de vie, et à ainsi distinguer son soi d'autrui et du monde externe. C'est en construisant des modèles génératifs profonds permettant de déterminer la cause des sensations ressenties que le bébé construit son modèle du soi et celui du monde. Cette construction a pour point d'origine le soi minimal, donc la corporalité. Selon Aikaterini Fotopoulou et Manos Tsakiris (2017),

in everyday life, feeding, sleeping, calming down, or entertaining routines typically include endless repetitions of multisensory bundles from at least two bodies (e.g. active and passive touch, proprioceptive and vestibular information, smell, temperature, visual and auditory feedback). During such experiences, the infant is therefore learning and responding to regularities and irregularities between the various sensorimotor “bundles” of his and the caregivers body. As the baby learns to bind together “own body” sensations from proprioception and vision regarding the position of her own limbs for example, she may also bind together sensations from proprioception, interpersonal touch and the smell of her father. (p. 13)

En d'autres mots, lorsque le parent (ou toute autre personne prenant soin du bébé) est en contact physique avec son bébé, ce dernier doit déterminer la cause de ses perceptions sensorielles. Si chaque fois que le bébé pleure et s'agite, il se fait allaiter, il comprend la boucle action-perception entre *son* action (pleurer, bouger ses membres, etc.), qui est au départ involontaire, et la perception sensorielle qui en résulte (goûter le lait maternel, se faire bercer, etc.). Une fois ce lien fait, c'est-à-dire une fois que la perception a informé le bébé sur ce qui cause cette dernière, celui-ci identifie ce qui est causé par lui versus ce qui est causé par autrui. Cela implique qu'il y a, de ce fait, distinction entre soi et autrui (à partir d'une représentation indifférenciée soi-autrui).

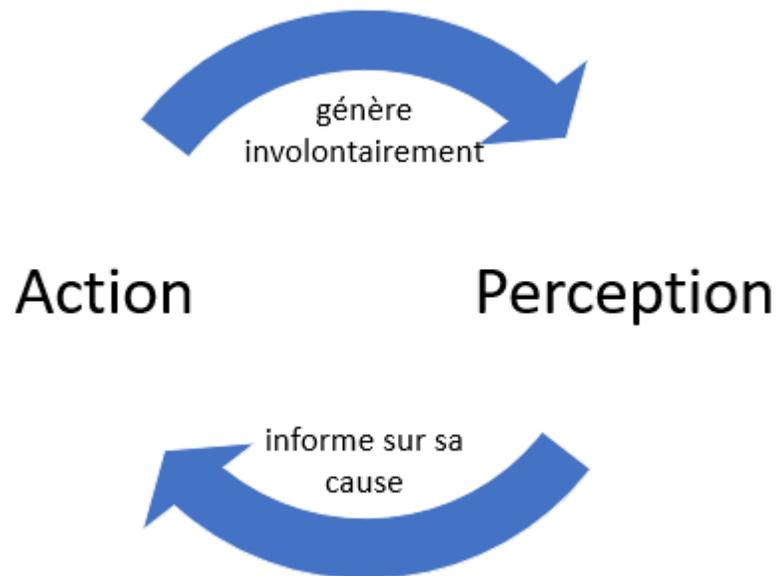


Figure 3.4 Boucle action-perception

Les illusions perceptuelles engendrant une identification erronée de son propre corps illustrent bien ce qui est ici avancé. Les expériences d'illusion de la main en caoutchouc, par exemple, montrent de quelle manière un individu dont les fonctions cérébrales sont tout à fait normales identifie une main en caoutchouc comme étant la sienne lorsque certaines circonstances expérimentales sont réunies. L'expérience en question, qui a été menée pour la première fois en 1998 par Matthew Botvinick et Jonathan Cohen²⁴, consiste à mettre la main gauche du participant hors de son champ de vision et à placer au lieu de celle-ci une main en caoutchouc mais aux allures réalistes. Il s'agit ensuite

²⁴ Cette expérience a été reproduite en 2004 par Ehrsson et al., puis en 2008 par Makin et al.

pour l'expérimentateur de caresser de manière synchrone à l'aide d'un pinceau à la fois la main en caoutchouc, qui est dans le champ de vision du participant, et sa main gauche, qui est hors de son champ de vision. Cela a pour résultat de donner l'impression au participant que la main en caoutchouc est la sienne. Il y a alors identification erronée d'une partie du corps du participant et « incorporation » du corps étranger en question dans la représentation qu'a le participant de son propre corps.

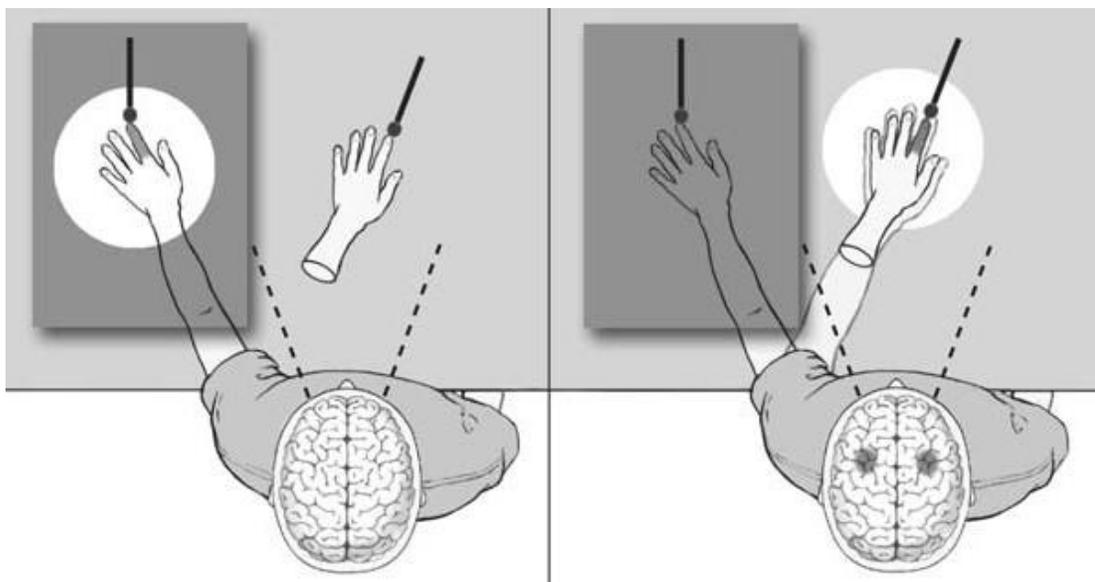


Figure 3.5 Tirée de *The Ego Tunnel: the science of the mind and the myth of the self* (Metzinger, 2009)

Cette erreur est due, selon le cadre explicatif du traitement prédictif, à la combinaison des croyances initiales (*priors*) *top-down* (la main gauche étant hors du champ de vision, la main en caoutchouc ayant pris sa place) et à l'*input bottom-up* (la perception sensorielle résultant de la caresse synchrone de la main gauche et de la main en caoutchouc) (de Vignemont et al., 2005; Tsakiris & Haggard, 2005; de Preester & Tsakiris, 2009; Hohwy & Paton, 2010; Limanowski & Blankenburg, 2013; Tsakiris, 2010). Dans ce cas-ci, la main en caoutchouc est la réponse la plus probable à la question « quelle est la cause de la donnée sensorielle ressentie? », en ce que

l'information sensorielle *bottom-up* qui est fournie à ce moment-là (caresse), combinée au jugement perceptuel sur la localisation de la main (*prior*) résultant de l'information visuelle obtenue identifie la main en caoutchouc comme faisant partie du corps (de Preester and Tsakiris, 2009; Limanowski & Blankenburg, 2013; Tsakiris, 2010). Cela démontre en quoi l'auto-identification et la modélisation de soi est un processus probabiliste : la compétition entre les croyances initiales (*priors*) et l'information sensorielle (*inputs*) ponctuelle donne lieu à une identification relative et graduelle de soi (Tsakiris & Haggard, 2005).

Une autre expérience permettant de montrer la nature hiérarchique et probabiliste du soi est celle de l'illusion du corps virtuel, telle que décrite dans *Video (Ergo Sum: Manipulating Bodily Self-Consciousness)* (Lenggenhager et al., 2007)²⁵. L'expérience en question consiste à disjoindre le sentiment de soi et la localisation de son propre corps chez des participants au fonctionnement cérébral tout à fait normal, de sorte à attribuer son soi à un corps qui n'est en fait que virtuel. Ce que cette expérience réussit à démontrer, c'est que la conscience de soi *corporelle* est ancrée dans le traitement *cognitif* et multisensoriel de l'information corporelle (Lenggenhager et al., 2007). Inspirée de l'expérience de l'illusion de la main de caoutchouc, l'illusion du corps virtuel traite de l'identification de soi via le sentiment de possession d'un corps dans son ensemble (*global ownership of the body as a whole*), par rapport à l'expérience de la main de caoutchouc qui, elle, traitait de l'identification de soi via le sentiment de possession de parties spécifiques du corps (*sense of body-part ownership*). Or, tel que discuté précédemment, le soi minimal phénoménal se caractérise entre autres par la capacité d'identification à son corps dans son ensemble : « Central defining features [the minimal phenomenal self] are, (i) a globalized form of identification with the body

²⁵ Cette expérience est décrite plus en détails dans la section suivante.

as a whole (as opposed to ownership for body parts)... » (Metzinger & Blanke, 2008, p. 8). Autrement dit, les représentations du corps dans son ensemble sont liées au sentiment de soi :

whole-body representations or global ownership are directly associated with the sense of selfhood. Studies on the [rubber-hand illusion] and somatoparaphrenia thus investigated only body-part ownership or the attribution and localization of a body part with respect to the global bodily self; i.e., a part-to-whole relationship. Thus, these studies did not experimentally manipulate selfhood per se. (Lenggenhager et al., 2007, p. 1097).

L'expérience de l'illusion du corps virtuel, qui s'opère en deux parties, traite donc du sentiment de soi et montre de quelle manière il est possible de désincarner celui-ci en l'attribuant à un autre corps qui lui est étranger et même inexistant. Dans la première partie de l'expérience (voir encadré A de la figure 3.7), le participant visualise à l'aide d'un casque de réalité virtuelle le dos de son propre corps (qui est filmé en temps réel à une distance de 2 mètres et projeté dans le casque en question dans une vidéo en trois dimensions). Pendant une durée d'une minute, le dos du participant est caressé à l'aide d'un outil, tantôt de manière synchrone, tantôt de manière asynchrone avec la vidéo projetée. Cette partie de l'expérience a eu pour résultat une forte identification avec le corps virtuel dans les cas de stimulation synchrone, contrairement aux cas de stimulation asynchrone, qui n'ont pas engendré une telle identification.

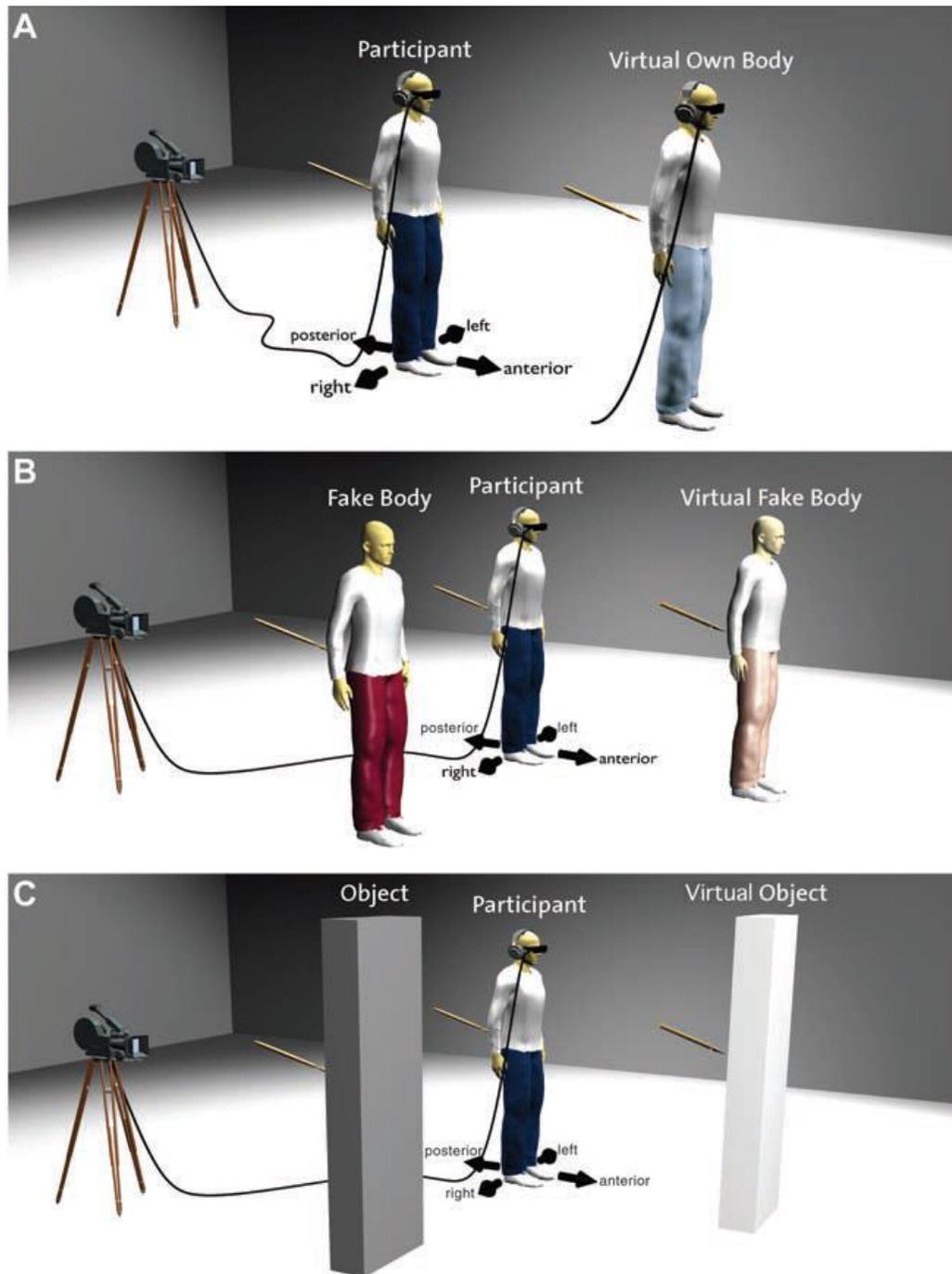


Figure 3.6 Tirée de *Video Ergo Sum: Manipulating Bodily Self-Consciousness* (Lenggenhager et al., 2007)

Dans la deuxième partie de l'expérience, l'image projetée dans le casque de réalité virtuelle du participant est soit la même que celle dans la première partie de l'expérience (encadré A de la figure 3.7), c'est-à-dire une vidéo filmée en temps réel du dos du participant, soit une vidéo filmée en temps réel du dos d'un mannequin (encadré B de la figure 3.7), soit une vidéo filmée en temps réel d'un objet non-corporel quelconque (encadré C de la figure 3.7). Dans les trois cas, et tout comme pour la première partie de l'expérience, la stimulation est tantôt synchrone, tantôt asynchrone. Cette deuxième partie de l'expérience a eu pour résultat une forte identification, dans les cas de stimulation synchrone, avec le corps virtuel ainsi que le mannequin, mais pas avec l'objet non-corporel. Ce résultat indique que la cognition joue un rôle important dans l'identification de soi – le fait de savoir que ce qui est stimulé est un objet non-corporel atténue l'identification avec celui-ci : « when asked whether it felt as if the virtual character or object was their body, participants self-identified with both bodily stimuli but not with the object during synchronous stroking. » (Lenggenhager et al., 2007, p. 1098). Nous pourrions également penser que ce résultat est dû au fait que l'objet non-corporel ne génère aucune des affordances²⁶ habituellement générées par le corps.

Ces exemples d'identification illusoire montrent en quoi la modélisation de soi et du monde est en fait une question de régularité environnementale et contextuelle. L'identification de soi est une inférence à la meilleure explication qui dépend de l'information disponible. Étant donné la régularité élevée des causes de nos états, nous ne faisons généralement pas d'erreurs quant à notre propre identification. Cette stabilité nous porte à croire, à tort, que notre soi est une entité, immuable et inaltérable sauf en

²⁶ Par « affordance » nous entendons « opportunités d'action » (Anderson & Chemero, 2009; Gibson, 1966)

cas de pathologie. Mais les expériences telles que celles qui ont ici été présentées démontrent que le soi est en fait un processus probabiliste, hiérarchique et inférentiel qui, dans certaines conditions (certes, hors du commun), peut être mal identifié, et ce, même chez des sujets ne souffrant d'aucune pathologie. C'est précisément cette équation qui donne lieu à l'identité personnelle et ce sont ces causes qui deviennent indissociables de soi. En effet, la capacité à constituer les causes endogènes en question repose sur la régularité et la récurrence des *inputs*. Ce principe s'appliquant à tous les niveaux hiérarchiques du soi, ce sont ces régularités et ces récurrences qui donneront lieu ultimement à *nos* désirs, croyances et à *notre* personnalité en général, ou autrement dit, à notre subjectivité. Nous pourrions alors revisiter et reconstruire en termes prédictivités la définition de Proust de la constitution du soi explorée dans le premier chapitre.

À la lumière de ce qui vient d'être discuté, on peut voir l'importance de définir de manière plus précise le concept d'identité personnelle. Comme nous l'avons précédemment affirmé, il est nécessaire d'avant tout déterminer de quelle partie du soi, ou de la personne, nous cherchons à établir l'identité. La corporalité comme critère de persistance de l'identité personnelle à travers le temps peut être tout à fait valable et défendable si c'est l'identité d'un niveau hiérarchique inférieur du soi que nous cherchons à déterminer. Nous pourrions en dire tout autant en ce qui a trait aux critères psychologiques de la persistance de l'identité personnelle à travers le temps : si c'est à l'identité de l'aspect métacognitif du sujet que l'on s'intéresse, alors ces critères sont tout à fait fondés. Nous pourrions dire qu'il y a autant de critères de persistance de l'identité personnelle à travers le temps qu'il y a de niveaux hiérarchiques – chercher à n'en identifier qu'un nierait la nature du soi, c'est-à-dire un processus graduel et hiérarchique et non une entité substantielle immuable. Avec un travail rigoureux de distinctions conceptuelles et à l'aide d'études expérimentales telles que celles de l'illusion de la main de caoutchouc et de l'illusion du corps virtuel, il y aurait moyen d'identifier de manière fiable et robuste des critères de persistance de l'identité

personnelle à travers le temps, mettant ainsi fin à l'impasse discursive qui a cours dans le cadre de ce débat. Il serait alors possible de contourner les oppositions théoriques qui n'ont pas lieu d'être et d'établir une réelle classification des critères fondée sur des données rendant compte de toute la complexité du soi et, par conséquent, de l'identité personnelle. Ce n'est toutefois pas dans ce mémoire que ces critères seront identifiés, une telle entreprise dépassant clairement le cadre de celui-ci – comme nous venons de l'affirmer, pour établir de manière rigoureuse des critères de persistance de l'identité personnelle à travers le temps, il est important de fonder cet examen sur un travail de distinction conceptuelle (de bien plus grande envergure que celle effectuée dans ce mémoire) et de valider les critères potentiels proposés grâce à des expériences empiriques. Un tel projet pourrait par exemple être l'objet d'un travail de recherche doctoral.

CONCLUSION

Dans ce mémoire, nous avons commencé par présenter les critères psychologique et corporel de persistance de l'identité personnelle à travers le temps, tels qu'ils ont été discutés par Derek Parfit (1984, 2008, 2012) et Judith Jarvis Thomson (2008). Nous avons montré qu'étant fondés sur des expériences de pensée, leur popularité et leur caractère intuitif est dû au fait qu'ils procèdent des perspectives à la première et à la deuxième personne, soit des perspectives subjective et intersubjective. Cela a pour effet un problème de circularité argumentaire. Dans ce premier chapitre, nous aurons donc montré que le problème de circularité dû au caractère intuitif et subjectif des critères psychologique et corporel tels que discutés par Parfit et Thomson nous empêche d'établir de manière objective leur validité et ne fait que confirmer nos propres intuitions.

Dans le deuxième chapitre, nous avons présenté des versions naturalisées de ces critères, à commencer par la redéfinition du critère lockéen par Joëlle Proust (2003), qui réussit à échapper au piège du problème de la circularité. Nous avons toutefois vu en quoi ce critère ne rend compte que d'un aspect de l'identité personnelle, ou du soi, car, comme nous l'aurons vu dans le premier chapitre, la validité d'un critère donné dépend du type d'identité que l'on cherche à déterminer, ce qui implique que plus d'un critère peut être valable. Cela nous aura amené à discuter de la théorie de l'esprit étendu (Clark & Chalmers, 1998), selon laquelle l'esprit s'étend au corps et à l'environnement, amenuisant ainsi l'opposition corps-esprit. En continuité avec cette idée, et en nous référant aux travaux de Colombetti (2013), Gallagher (2005) et Legrand (2007), nous avons présenté la notion de pré-réflexivité corporelle, soit la forme la plus élémentaire de conscience phénoménale de soi. Cela aura eu pour objectif d'établir les bases pour

la suite de notre argumentaire, soit que le soi est hiérarchique et qu'à son fondement se trouve la corporalité.

Dans le troisième et dernier chapitre, nous présentons de manière plus détaillée la nature hiérarchique du soi. Construisant sur l'aspect pré-réflexif du soi, nous avons présenté la classification hiérarchique du soi selon Metzinger (2003, 2008, 2009, 2010, 2013b, 2017). À la base de cette hiérarchie se trouve le soi minimal phénoménal et à son niveau supérieur se trouve le soi cognitif. Nous avons montré grâce à cette classification hiérarchique que la dichotomie corps-esprit n'a pas lieu d'être, sachant que l'expérience du corps et de l'esprit font tous deux partie de la constitution du soi et qu'ils sont interdépendants. Outre le caractère hiérarchique du soi, nous avons également discuté du caractère probabiliste du soi, en présentant la théorie du traitement prédictif (Clark, 2013, 2016; Hohwy, 2013, 2016; Seth, 2013; Wiese & Metzinger, 2017). Cela nous a permis de montrer que les différents niveaux du soi reposent sur des inférences hiérarchiques et que l'auto-identification est en fait relative et graduelle. Pour en faire la démonstration, nous avons présenté deux expériences: l'illusion de la main en caoutchouc et l'illusion du corps virtuel. La première expérience montre comment, dans certaines circonstances, il est possible d'« intégrer » un corps étranger à son propre corps et la seconde montre qu'il est même possible d'attribuer son soi à une entité totalement fictive mais ressemblant en tous points à son propre corps.

Comme nous l'avons vu tout au long de ce mémoire, la notion d'identité dont il est ici question est à distinguer de celle du sens commun. Il ne s'agit pas de faire appel à notre intuition pour déterminer ce que nous croyons qui fait de nous les mêmes à travers le temps (cette manière de raisonner donnant lieu à un problème de circularité, tel que précédemment établi). Il s'agit plutôt d'utiliser les outils de réflexion que nous offrent les théories en philosophie des neurosciences cognitives (telles que celle du modèle hiérarchique du soi et celle du traitement prédictif) ainsi que les expériences empiriques

permettant de valider ou non nos intuitions et de confirmer ou non les théories à l'usage dans l'étude des problèmes théoriques comme celui de la persistance de l'identité personnelle (telles que celle de l'illusion de la main en caoutchouc et celle de l'illusion du corps virtuel) pour stimuler notre réflexion et enrichir notre appareil conceptuel. L'usage de tels outils pourrait même avoir pour effet de changer nos intuitions, de manière à ajouter à notre perspective subjective et indépassable une couche interprétative pouvant nous apporter une nouvelle perspective sur des questions fondamentales.

Selon le principe de la formation virtuelle de l'identité, tel qu'élaboré par Metzinger (2013, p. 5), l'identité personnelle serait une conséquence d'un modèle du soi capable de s'autosimuler comme étant le même à travers le temps, ce qui permet des formes adaptatives de cognition de haut niveau, dont la cognition sociale et morale. En effet, un organisme complexe capable de s'autosimuler comme étant le même à travers le temps peut se sentir concerné par des récompenses ou sanctions futures, ce qui est nécessaire à la cognition sociale et morale. Ainsi, l'identité personnelle à travers le temps serait une illusion générée par un modèle du soi virtuel capable de générer des autotromperies narratives qui sélectionnent des objectifs futurs en fonction des expériences passées. Suivant cette conception de l'identité personnelle, il n'y aurait alors aucun critère précis qui permettrait de fonder une identité personnelle à travers le temps (ou une quelconque forme de soi, puisque celui-ci est un processus hiérarchiquement modélisé) (Metzinger, 2013).

Dans cette perspective, le caractère intuitif des différents critères de l'identité personnelle à travers le temps émerge des contraintes fonctionnelles du modèle du soi phénoménalement transparent. L'intuitivité du critère corporel pourrait émerger de l'auto-identification avec un modèle de soi transparent capable d'agentivité corporelle. En ce sens, le soi corporel s'impose comme une réalité en vertu de la transparence phénoménale, qui force l'identification avec le contenu du modèle du soi minimal

phénoménal, de sorte que l'agent en question est capable de contrôler son propre corps, ce qui serait nécessaire pour garantir la survie de l'organisme. En effet, une perte de transparence du modèle du soi minimal phénoménal pourrait avoir des conséquences dommageables pour l'identité personnelle, telles que des symptômes associés au trouble de dépersonnalisation (Limanowski & Blankenburg, 2013). L'intuitivité du critère psychologique pourrait émerger de l'auto-identification avec un modèle de soi transparent capable d'agentivité cognitive. En effet, le soi cognitif s'impose comme une réalité en vertu de la transparence phénoménale, qui implique l'auto-identification avec le contenu d'un modèle de l'agent épistémique, capable de contrôler ses pensées (ou autotromperies narratives automatiques, qui fournissent l'illusion du soi qui persiste à travers le temps). En ce sens, le soi psychologique s'impose comme une réalité en vertu de la transparence phénoménale, qui force l'auto-identification avec le contenu d'un modèle de l'agent épistémique capable de contrôler la pensée, ce qui serait nécessaire pour garantir la survie de l'organisme. Là encore, la perte de transparence phénoménale du modèle de l'agent épistémique pourrait entraîner des conséquences dommageables pour l'identité personnelle, telles que les délires d'insertion de pensée dans les troubles schizophréniques (Metzinger, 2013). Quoi qu'il en soit, les critères de l'identité personnelle seraient infondés selon Metzinger puisque le soi est un processus de modélisation hiérarchique qui repose sur l'illusion subjective convaincante d'un soi corporel ou psychologique substantiel afin de permettre la survie de l'organisme.

BIBLIOGRAPHIE

- Anderson, M. L. & Chemero, A. (2009). Affordances and intentionality: Reply to Roberts. *Journal of Mind and Behavior*, 30(4), 301-312.
- Ayer, A. J. (1936). *Language, Truth, and Logic*. London : Gollancz.
- Ayers, M. (1990). *Locke*, vol. 2. London : Routledge.
- Botvinick, M., Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, 391, 756. <https://doi.org/10.1038/35784>
- Butler, J. (1842). *The Analogy of Religion, Natural and Revealed, to the Constitution and Course of Nature*. London: James, John, and Paul Knapton.
- Burn, D. J. & Trösters, A. I. (2004). Neuropsychiatric complications of medical and surgical therapies for Parkinson's disease. *Journal of Geriatric Psychiatry and Neurology*, 17(3), 172-180. <https://doi.org/10.1177/0891988704267466>
- Carter, W. R. (1989) How to Change Your Mind. *Canadian Journal of Philosophy*, 19. 1-14.
- Clark, A. & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7-19. <https://doi.org/10.1093/analys/58.1.7>
- Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford : Oxford University Press.
- Colombetti, G. (2013). *The Feeling Body: Affective Science Meets the Enactive Mind*. Cambridge : MIT Press.
- Cummins, R. (1998). Reflections on Reflective Equilibrium. Dans M. DePaul et W. Ramsey (dir.), *Rethinking Intuition* (113-128). New York : Rowman & Littlefield.

- De Preester, H., & Tsakiris, M. (2009). Body-extension versus body-incorporation: is there a need for a body-model? *Phenomenology and the Cognitive Sciences*, 8, 307-319. <https://doi.org/10.1007/s11097-009-9121-y>
- De Vignemont, F., Tsakiris, M., & Haggard, P. (2005). Body mereology. Dans G. Knoblich, I. M. Thornton, M. Grosjean & M. Shiffrar (dir.), *Human Body Perception from the Inside Out* (147-170). New York : Oxford University Press.
- Dennett, D. (1980). The milk of human intentionality. *Behavioral and Brain Sciences*, 3(3), 428-430. <https://doi.org/10.1017%2Fs0140525x0000580x>
- Dennett, D. (1991). *Consciousness Explained*. London : Allen Lane.
- Dehaene, S. (2018). Psychologie cognitive expérimentale. *L'annuaire du Collège de France*, 116, 183-195. <https://doi.org/10.4000/annuaire-cdf.12823>
- Ehrsson H. H., Spence C., Passingham R. E. (2004) That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science*, 305(5685), 875-877. <https://doi.org/10.1126/science.1097011>
- Evans, G. (1982). *The Varieties of Reference*. Oxford : Oxford University Press.
- Friston, K. J. (2017). Self-evidencing babies: Commentary on "Mentalizing Homeostasis: The Social Origins of Interoceptive Inference" by Fotopoulou & Tsakiris. *Neuropsychoanalysis*, 19(1), 43-47. <https://doi.org/10.1080/15294145.2017.1295216>
- Fotopoulou, A. & Tsakiris, M. (2017). Mentalizing homeostasis: The social origins of interoceptive inference, *Neuropsychoanalysis*, 19(1), 3-28. <https://doi.org/10.1080/15294145.2017.1294031>
- Frith, C. (2007). *Making up the Mind: How the Brain Creates our Mental World*. Oxford : Blackwell Publishing Ltd.
- Gallagher, S. (2005). *How the body shapes the mind*. Oxford : Oxford University Press.
- Garrett, B. (1998). *Personal Identity and Self-Consciousness*. London : Routledge.

- Gibson, J.J. (1966). *The senses considered as perceptual systems*. Boston : Houghton Mifflin.
- Hohwy, J., & Paton, B. (2010). Explaining away the body: experiences of supernaturally caused touch and touch on non-hand objects within the rubber hand illusion. *PLoS One*, 5(2), e9416. <https://doi.org/10.1371/journal.pone.0009416>
- Hohwy, J. (2013). *The Predictive Mind*. Oxford : Oxford University Press.
- Hohwy, J. (2016). The Self-Evidencing Brain. *Noûs*, 50(2), 259-285.
- Hohwy, J. & Michael, J. (2017). Why Should Any Body Have a Self? Dans F. De Vignemont et A. J. T. Alsmith (dir.), *The Subject's Matter: Self-Consciousness and the Body* (pp. 363-391). Cambridge : The MIT Press.
- Hudson, H. (2001). *A Materialist Metaphysics of the Human Person*. Ithaca : Cornell University Press.
- Hudson, H. (2007). I Am Not an Animal!. Dans P. van Inwagen et D. Zimmerman (dir.), *Persons: Human and Divine* (216-234). Oxford : Oxford University Press.
- James, W. (1918). *The Principles of Psychology*. New York : Henry Holt and Company.
- Johnston, M. (1987). Human Beings. *Journal of Philosophy*, 84(2), 5-83.
- Johnston, M. (2016). Remnant Persons: Animalism's Undoing. Dans S. Blatti et P. Snowdon (dir.), *Animalism: New Essays on Persons, Animals, and Identity* (89-127). Oxford : Oxford University Press.
- Knobe J. & Nichols, S. (dir.) (2008). *Experimental Philosophy*. New York : Oxford University Press.
- Kornblith, H. (2002). *Knowledge and Its Place in Nature*. Oxford : Oxford University Press.
- Langford, S. (2017). A Defence of Anti-Criterialism. *Canadian Journal of Philosophy*, 47, 613-630.

- Legrand, D. (2007). Pre-Reflective Self-Consciousness: On Being Bodily in the World. *Janus Head*, 9(2), 493-519.
- Lenggenhager, B., Tadi, T., Metzinger, T. & Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science*, 317(5841), 1096-1099.
<https://doi.org/10.1126/science.1143439>
- Lewis, D. (1976). Survival and Identity. Dans A. Rorty (dir.), *The Identities of Persons* (p. 17-40). Berkeley : University of California Press.
- Limanowski, J. & Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Frontiers in Human Neurosciences*, 7, 547.
<https://doi.org/10.3389/fnhum.2013.00547>
- Lowe, E. J. (1996) *Subjects of Experience*. Cambridge : Cambridge University Press.
- Lowe, E. J. (2012) The Probable Simplicity of Personal Identity. Dans G. Gasser & M. Stefan (dir.), *Personal Identity: Simple or Complex?*. Cambridge : Cambridge University Press.
- Locke, J. (2008). *An Essay concerning Human Understanding*. Oxford : Oxford University Press.
- Machery, E. (2009). *Doing Without Concepts*. Oxford : Oxford University Press.
- Mackie, D. (1999). Personal Identity and Dead People, *Philosophical Studies*, 95. 219-242.
- Makin, T. R., Holmes, N. P., & Ehrsson, H. H. (2008). On the other hand: dummy hands and peripersonal space. *Behavioural Brain Research*, 191, 1–10.
<https://doi.org/10.1016/j.bbr.2008.02.041>
- Merricks, T. (1998). There Are No Criteria of Identity Over Time. *Noûs*, 32, 106-124.
- Metzinger, T. (2003). *Being no one: The self-model theory of subjectivity*. Cambridge : MIT Press.
- Metzinger, T. & Blanke, O. (2008). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, 13(1), 7-13.

- Metzinger, T. (2009). *The ego tunnel: The science of the mind and the myth of the self*. New York : Basic Books.
- Metzinger, T. (2010). The No-Self-Alternative. Dans S. Gallagher (dir.), *Oxford Handbook of the Self* (p. 279-296). Oxford : Oxford University Press.
- Metzinger, T. (2013). The myth of cognitive agency: subpersonal thinking as a cyclically recurring loss of mental autonomy. *Frontiers in Psychology*, 4(746), 1-17. <https://doi.org/10.3389/fpsyg.2013.00746>
- Metzinger, T. (2013b). Why are dreams interesting for philosophers? The example of minimal phenomenal selfhood, plus an agenda for future research. *Frontiers in Psychology*, 4(1), 1-17. <https://doi.org/10.3389/fpsyg.2013.00746>
- Metzinger, T. (2016). Suffering. Dans K. Almqvist & A. Haag (dir.), *The Return of Consciousness: A new science on old questions* (221-248). Stockholm : Axel and Margaret Ax:son Johnson Foundation.
- Metzinger, T. (2017). The Problem of Mental Action - Predictive Control without Sensory Sheets. Dans T. Metzinger & W. Wiese (dir.), *Philosophy and Predictive Processing* (1-26). Frankfurt am Main: MIND Group.
- Metzinger, T. (2017b). Why is Mind Wandering Interesting for Philosophers?
Récupéré de
https://www.blogs.uni-mainz.de/fb05philosophieengl/files/2017/12/Metzinger_Mind_Wandering_OHST_2017_penultimate_draft.pdf
- Metzinger, T. & Wiese, W. (2017). Vanilla PP for Philosophers: A Primer on Predictive Processing. Dans T. Metzinger & W. Wiese (dir.), *Philosophy and Predictive Processing* (1-18). Frankfurt am Main: MIND Group.
- Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, 83(4), 435-450.
- Nagel, T. (1986). *The View from Nowhere*. New York : Oxford University Press.
- Nichols, S. & Bruno, M. (2010). Intuitions about Personal Identity: An Empirical Study. *Philosophical Psychology*, 23, 293-312.
- Noonan, H. (2003). *Personal Identity* (2e ed.). London : Routledge.

- Olson, E. (1997). *The Human Animal: Personal Identity Without Psychology*. New York : Oxford University Press.
- Olson, E. T. (2019). Personal Identity. Dans *The Stanford Encyclopedia of Philosophy*. Récupéré de <https://plato.stanford.edu/archives/fall2019/entries/identity-personal/>
- Olson, E. T. (2021). Personal Identity. Dans *The Stanford Encyclopedia of Philosophy*. Récupéré de <https://plato.stanford.edu/archives/spr2021/entries/identity-personal/>
- Parfit, D. (1971). Personal Identity. *Philosophical Review*, 80(1), 3-27.
- Parfit, D. (1984). *Reasons and Persons* [Format ePub]. New York: Oxford University Press.
- Parfit, D. (2008). Persons, Bodies, and Human Beings. Dans T. Sider, J. Hawthorne et D. W. Zimmerman (dir.), *Contemporary Debates in Metaphysics* (p. 177-208). London : Blackwell Publishing.
- Parfit, D. (2012). We Are Not Human Beings. *Philosophy*, 87(1), 5-28.
- Perry, J. (1972). Can the Self Divide? *Journal of Philosophy*, 69(16), 463-488.
- Proust, J. (2003). Thinking of oneself as the same. *Consciousness and Cognition*, 12(4), 495-509. [https://doi.org/10.1016/S1053-8100\(03\)00077-1](https://doi.org/10.1016/S1053-8100(03)00077-1)
- Pust, J. (2000). *Intuitions as Evidence*. New York : Garland/Routledge.
- Ricœur, P. (1990). *Soi-même comme un autre*. Paris : Seuil.
- Sartre, J.-P. (1943). *L'être et le néant*. Paris : Gallimard.
- Schechtman, M. (1996). *The Constitution of Selves*. Ithaca : Cornell University Press.
- Schroer, J. W. & Schroer, R. (2014). Getting the Story Right: A Reductionist Narrative Account of Personal Identity. *Philosophical Studies*, 171, 445-469.
- Seigel, J. (2005). *The Idea of the Self: Thought and Experience in Western Europe Since the Seventeenth Century*. Cambridge : Cambridge University Press.

- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Cognitive Sciences*, 17(11), 656–63.
- Shoemaker, S. (1970). Persons and Their Pasts. *American Philosophical Quarterly*, 7(4) 269–285.
- Shoemaker, S. (1984). Personal identity: A materialist's account. Dans S. Shoemaker et R. G. Swinburne (dir.), *Personal identity* (67-138). Oxford : Blackwell Publishing.
- Shoemaker, S. (1997). Self and Substance. Dans J. Tomberlin (dir.), *Philosophical Perspectives 11: Mind, Causation, and World* (283–319). Atascadero : Ridgeview Publishing.
- Shoemaker, S. (1999). Self, Body, and Coincidence. *Proceedings of the Aristotelian Society*, 73(1), 287-306. <https://doi.org/10.1111/1467-8349.00059>
- Shoemaker, S. (2008). Persons, Animals, and Identity. *Synthese*, 163, 313-324. <https://doi.org/10.1007/s11229-007-9253-y>
- Shoemaker, S. (2011). On What We Are. Dans S. Gallagher (dir.). *The Oxford Handbook of the Self* (352-371). Oxford : Oxford University Press.
- Swinburne, R. (1984) Personal Identity: The Dualist Theory. Dans S. Shoemaker & R. Swinburne (dir.). *Personal Identity* (1-22). Oxford: Blackwell.
- Taylor, C. (1985). Self-interpreting Animals. Dans *Human Agency and Language: Philosophical Papers 1* (97-114). Cambridge : Cambridge University Press.
- Thompson, E., Lutz, A. & Cosmelli, D. (2005). Neurophenomenology: An introduction for neurophilosophers. Dans A. Brook et K. Akins (dir.), *Cognition and the Brain: The Philosophy and Neuroscience Movement* (40-97). Cambridge : Cambridge University Press.
- Thomson, J. J. (2008). People and Their Bodies. Dans T. Sider, J. Hawthorne et D. W. Zimmerman (dir.), *Contemporary Debates in Metaphysics* (p. 155-175). London : Blackwell Publishing.
- Tsakiris, M., & Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 80-91. <https://doi.org/10.1037/0096-1523.31.1.80>

Tsakiris, M. (2010). My body in the brain: a neurocognitive model of body-ownership. *Neuropsychologia*, 48, 703-712.

<https://doi.org/10.1016/j.neuropsychologia.2009.09.034>

Unger, P. (1990). *Identity, Consciousness, and Value*. Oxford : Oxford University Press.

Unger, P. (2000). The Survival of the Sentient. Dans J. E. Tomberlin (dir.), *Philosophical Perspectives 14: Action and Freedom* (325-348). Malden: Blackwell Publishing.

Van Inwagen, P. (1990). *Material Beings*. Ithaca : Cornell University Press

Williams, B. (1956). Personal Identity and Individuation. *Proceedings of the Aristotelian Society*, 57, 229-252.

Williams, B. (1970). The Self and the Future. *Philosophical Review*, 79(2), 161-180.

Wollos, J. (2016). *That Seems Right: Reasoning, Inference, and the Feeling of Correctness* (Thèse de doctorat). Université Columbia. Récupéré de <https://academiccommons.columbia.edu/doi/10.7916/D81J9B1V>