L'INTROSPECTION, APPROCHE PHILOSOPHIQUE

Par ***Pascal Ludwig*** et ***Matthias Michel*****,** Université de Paris-Sorbonne

Résumé

L'introspection est traditionnellement conçue comme une source de connaissance qui possède un certain nombre de caractéristiques : elle semble garantir une voie d'accès privilégiée et immédiate à nos propres états et processus mentaux, conférant au sujet des connaissances sûres et une autorité spéciale, en première personne, sur les connaissances ainsi acquises. Différentes approches se sont développées en philosophie, dans un débat encore ouvert, pour rendre compte de l'introspection et du statut des connaissances qu'elle confère, que l'on peut toutes situer relativement au modèle cartésien de l'accès au domaine du mental.

Table des matières

1. Introduction

Supposons qu'un agent, Hoeradip, observe un pic vert en train de chercher de la nourriture dans le jardin sur lequel donne son bureau. Cette expérience lui apporte des connaissances sur le monde : sur la forme et la couleur de l'oiseau, les caractéristiques de son cri, les trajectoires de ses mouvements. Mais elle est aussi l'occasion de former des connaissances non pas sur les objets extérieurs, mais bien plutôt sur les caractéristiques de ses propres états mentaux. Cela implique que l'expérience de l'oiseau lui apprend quelque chose sur lui-même,

et pas seulement quelque chose sur le monde : par exemple, qu'il est en train d'éprouver telle ou telle sorte de sensation, et que cette expérience peut être comparée avec telle autre expérience qu'il a eu dans le passé et dont il se souvient. Nous désignerons ce type de connaissance de soi par l'expression « introspection ». L'introspection peut porter sur les perceptions, les sensations corporelles, les émotions, sur les jugements ou les décisions, mais également sur des états dits « dispositionnels » comme les croyances ou les désirs. Certains phénomènes psychologiques sont d'abord connus par introspection. Ainsi, nous avons tous un accès introspectif aux discours intérieurs que nous nous tenons à nous-mêmes, nous permettant de saisir les contenus des énoncés qui constituent ces monologues. Il est difficile de savoir si une personne dont on observe le comportement est en train de se parler à elle-même ; en revanche, il paraîtrait très étrange de se demander à soi-même si, à un certain moment, l'on est ou non en train de se livrer à cette action mentale, précisément parce que ce type d'épisode psychologique est facilement accessible par introspection. On ne peut pas plus ignorer que l'on se parle intérieurement qu'ignorer que l'on est en train de lire à voix basse. De même, il est sans doute beaucoup plus facile de savoir, par introspection, que l'on est en train d'éprouver une émotion complexe comme la nostalgie, que d'identifier un tel état d'esprit chez autrui en observant son comportement. Pour autant, il est probable que la grande majorité des états mentaux accessibles par introspection, si ce n'est pas tous, peuvent aussi être connus par d'autres méthodes. Cela est vrai aussi bien lorsqu'on considère les états mentaux des autres personnes qu'à propos de ses propres états : un agent peut découvrir qu'il est déprimé en observant l'image que lui renvoie un miroir, ou découvrir qu'il a envie de s'acheter une nouvelle voiture en constatant qu'il consulte de façon frénétique des magazines automobiles. Les connaissances introspectives appartiennent indubitablement à une classe qui les englobe : la classe des connaissances psychologiques portant sur nos propres états mentaux.

On le voit, l'ensemble de notre savoir portant sur nos propres états psychologiques ne provient pas de l'introspection. Il ne faut pas, autrement dit, identifier introspection et connaissance de soi. La question de la connaissance de soi déborde en effet très largement celle de l'introspection. On peut se connaître soi-même autrement que par introspection, et la connaissance de soi ne porte pas seulement sur les états psychologiques au sens strict : elle porte aussi sur nos traits de caractères, sur nos habitudes, sur nos capacités, sur nos vertus et nos vices moraux ou intellectuels, sur nos valeurs, etc. Dans la littérature contemporaine, la question de la nature de l'introspection et celle de la nature de la connaissance de soi ont tendance à se confondre, mais c'est sans doute dommage (Cassam 2014). Supposons, pour reprendre un exemple développé par Quassim Cassam que vous vous demandiez si vous êtes ou non raciste. Si vous parvenez à acquérir des croyances justifiées sur cette question, vous aurez sans doute appris à mieux vous connaître. Mais il est loin d'être clair qu'on puisse répondre à une telle question par introspection. Non pas que l'introspection ne puisse jouer aucun rôle dans la justification de la réponse. Par exemple, si vous réalisez, en réfléchissant introspectivement à vos croyances, que vous considérez qu'il existe des hiérarchies entre les groupes ethniques, vous pourrez certainement conclure que vous êtes racistes. Imaginons néanmoins que vous n'ayez aucune croyance explicitement raciste, mais que vous vous aperceviez que vous avez des réticences à développer des relations avec les membres d'autres groupes ethniques que le vôtre, que vous avez fait des efforts importants pour habiter dans un quartier ethniquement homogène, que vous ayez systématiquement tendance à sous-évaluer les compétences de vos collaborateurs lorsqu'ils appartiennent à une autre ethnie

que la vôtre, etc. Il semble que dans une telle situation vous pourriez conclure que vous êtes bel et bien raciste, sans avoir uniquement utilisé l'introspection. Ce qui montre bien que lorsqu'on parvient à mieux se connaître, y compris à mieux connaître ses propres dispositions psychologiques, ce n'est pas toujours ni uniquement grâce à l'introspection.

Existe-t-il des critères qui permettraient de distinguer clairement les connaissances psychologiques acquises par introspection de connaissances non-introspectives, mais portant sur le même domaine, autrement dit des critères permettant de définir la connaissance introspective, par opposition aux autres formes de la connaissance de soi ? Il est important de préciser d'emblée que ce qui nous intéresse n'est pas tant le sens du terme « introspection » dans le langage courant, mais plutôt le sens qu'il a pris dans la tradition philosophique et qu'il a encore aujourd'hui dans les discussions philosophiques. Il nous semble à cet égard pertinent de remonter à l'œuvre de Descartes, qui peut être considérée comme une des sources les plus importantes du concept philosophique d'introspection.

2. Introspection et modèle cartésien de l'accès au mental

Descartes n'emploie pas l'expression « introspection », mais il est sans doute à l'origine de l'importance accordée à cette méthode d'acquisition de connaissances sur soi en épistémologie. L'argument dit du « Cogito » a la structure suivante (Peacocke 2012 ; 2014) :

(1) Un épisode mental conscient ayant un contenu C, par exemple un épisode consistant à douter consciemment de la vérité d'une certaine proposition ;

(2) L'auto-attribution par le sujet de cet épisode, par exemple "je doute que C" ;

(3) Un jugement affirmant l'existence du sujet, "j'existe".

Le jugement (2) est introspectif au sens qui nous intéresse. Il s'agit d'un jugement qui exprime une connaissance de soi, sous la forme d'une auto-attribution, et d'une connaissance de soi qui porte sur un état mental. Par ailleurs la connaissance exprimée par (2) possède trois caractéristiques essentielles qui la distinguent d'autres sortes de connaissances de soi : elle est immédiate, elle possède une sécurité particulièrement élevée, et enfin elle est subjective, ce qui se manifeste par l'usage de la première personne dans (2). Examinons successivement ces trois caractéristiques.

1. L'immédiateté de la connaissance introspective

En premier lieu, les connaissances introspectives se caractérisent par leur immédiateté. Attribuer des états mentaux à autrui suppose, en général, l'exercice de capacités inférentielles. Revenons à l'exemple mentionné plus haut d'Hoeradip observant un pic vert. Dans un tel contexte, on peut inférer de son comportement qu'il regarde le pic vert, et qu'en conséquence son esprit se trouve dans certains états sensoriels conscients. Il s'agit cependant bien d'une inférence et pas d'une connaissance immédiate : si en observant mieux la situation on observe qu'Hoeradip a les yeux fermés, ou qu'il dirige son regard ailleurs que vers le pic vert, on sera amené à réviser la première hypothèse qui avait été formulée. En revanche, Hoeradip lui-même ne semble pas avoir besoin de raisonner pour savoir qu'il a une expérience consciente d'une certaine sorte lorsqu'il est en train d'avoir cette expérience. Descartes

insiste dans plusieurs textes sur l'immédiateté de la connaissance introspective qui sert de prémisse au Cogito. Comparons ainsi les deux inférences suivantes :

(I1) Je doute. J'existe.

(I2) Je marche. J'existe.

Hobbes, dans ses Troisièmes objections aux méditations métaphysiques, fait remarquer que (I2) est valide tout autant que (I1). Il a raison, mais cela ne remet pas en cause l'argument de Descartes. En effet, le statut épistémique de (I1) n'est pas le même que celui de (I2), car les prémisses sont justifiées, dans ces deux inférences, par des raisons complètement différentes. Je sais que je marche, selon Descartes, parce que je m'observe marchant et que j'infère de cette observation que j'ai la propriété de marcher (AT IX, 28). En revanche, le jugement introspectif « je doute » n'est pas justifié en vertu d'une inférence à partir d'une observation : l'expérience consciente du doute la justifie immédiatement, d'une manière qui n'est pas inférentielle. Descartes, dans l'appendice More geometrico aux Réponses aux secondes objections, définit d'ailleurs la pensée comme ce qui est immédiatement connaissable : « Pensée. J'utilise ce terme pour désigner tout ce qui est en nous, de sorte que nous en soyons immédiatement conscients » (AT VII, 160, nous soulignons).

1. **La sûreté de l'introspection**

En second lieu, l'introspection nous apporte des connaissances qui semblent particulièrement sûres. Le titre de la seconde méditation métaphysique de Descartes affirme ainsi que l'esprit humain est « plus aisé à connaître que le corps » (AT VII, 23). Ce qui semble clair, c'est que certaines connaissances introspectives au moins possèdent un degré de sûreté plus grand que les connaissances perceptives, c'est-à-dire qu'elles nous apportent des raisons de croire plus difficiles à remettre en question. Considérons ainsi une illusion visuelle, par exemple celle causée par l'image suivante :

L'expérience visuelle associée à cette image nous conduit à formuler deux jugements :

(4) Cette image représente des spirales colorées différemment (jugement perceptif).

(5) En voyant cette image j'ai une expérience visuelle de spirales colorées différemment (jugement introspectif).

Lorsqu'on agrandit l'image, on s'aperçoit qu'il s'agit d'une illusion, et que les spirales qui apparaissent comme vertes et bleues sont en réalité de la même couleur. En conséquence, le jugement de perception (4) doit être remis en question. Mais ce n'est pas le cas du jugement introspectif (5) : le fait que les spirales ne soient pas réellement de couleurs différentes ne nous empêche pas d'avoir une expérience visuelle exactement semblable à celle que nous aurions si nous percevions vraiment des spirales de couleurs différentes, et il n'y a donc pas de raison de remettre (5) en question. De façon plus générale, il semble que la distinction entre la façon dont les choses sont réellement, et la façon dont les choses apparaissent, n'ait pas d'application dans le domaine de la connaissance introspective. Dans l'exemple que nous venons de discuter, constater que l'expérience visuelle ne nous présente que des apparences de couleurs et non des couleurs réelles est une bonne raison de réviser le jugement (4). Mais

on voit mal, de prime abord du moins, comment l'introspection pourrait nous présenter les apparences des états mentaux plutôt que leur réalité. L'exemple des sensations de douleur est souvent pris pour illustrer ce point, et de nombreux philosophes soulignent qu'éprouver une sensation de douleur, c'est déjà avoir mal, et qu'une simple apparence de douleur est donc inconcevable (Kripke 1980 ; Dretske 2005 ; McGinn 1982). Parce que l'apparence d'états mentaux appréhendés de manière introspective n'est autre que la réalité de ces états, au moins si l'on suit l'intuition de Descartes, l'introspection semble donc constituer une méthode particulièrement sûre d'acquisition de connaissances, plus sûre que la perception. C'est clairement la position de Descartes, qui considère que les jugements introspectifs sont complètement immunisés contre le doute lorsqu'ils portent sur les contenus de nos expériences conscientes vécues en première personne, et qu'ils sont justifiés de façon certaine par ces expériences. Ainsi écrit-il que « je suis le même qui sens, c'est-à-dire qui reçoit et connaît les choses comme par les organes des sens, puisqu'en effet je vois la lumière, j'entends le bruit, je ressens la chaleur. Mais l'on dira que ces apparences sont fausses et que je dors. Qu'il en soit ainsi ; toutefois, à tout le moins, il est très certain qu'il me semble que je vois, que j'entends, et que je m'échauffe ; et c'est proprement ce qui en moi s'appelle sentir » (AT VII, 29 ; AT IX, 23 ; nous soulignons). Les auto-attributions « il me semble que je vois, que j'entends, et que je m'échauffe » sont introspectives, et donc plus sûres que des jugements perceptifs. Descartes les considère comme immunisées au doute.

1. **L'introspection et la perspective en première personne**

Nous l'avons vu, la connaissance introspective est un type de connaissance de soi. Mais il y a plus : c'est une sorte de connaissance de soi à laquelle on a accès précisément parce qu'on se trouve être soi. Lorsqu'Hoeradip, après avoir enlevé ses lunettes, se rend compte qu'il voit le pic vert qui se trouve devant lui de façon floue, et donc que son expérience visuelle actuelle diffère sensiblement de l'expérience visuelle qu'il avait avant de les retirer, il acquiert une connaissance sur lui-même d'une façon qui n'est accessible qu'à lui. On peut parler, à cet égard, d'un privilège de la première personne en ce qui concerne la connaissance introspective (Bar-On 2004). Cela ne signifie pas qu'un observateur extérieur n'ait aucun accès au contenu des états introspectifs du sujet. Lorsqu'Hoeradip perçoit le pic vert s'envoler, il a la sensation d'un mouvement dans son champ visuel, et cette sensation correspond à une activité dans l'aire MT de son cortex visuel. Il est possible pour un observateur d'observer cette activité de MT et d'en déduire qu'une sensation de mouvement est présente dans l'esprit d'Hoeradip (Naselaris et al. 2011). Néanmoins, il est clair que les méthodes par lesquelles Hoeradip et l'observateur extérieur arrivent à la conclusion selon laquelle une sensation de mouvement est présente diffèrent du tout au tout, et cela semble conférer une autorité épistémique particulière au sujet lorsqu'il est question de ses propres états mentaux (Shoemaker 1996). S'il est en général possible de remettre en question les jugements d'une personne sincère, à condition bien entendu d'avoir de bonnes raisons pour cela, cela semble inapproprié dans le cas des jugements introspectifs. Ainsi je peux faire remarquer à un interlocuteur que le jugement (4) mentionné plus haut n'est pas justifié, et lui expliquer pourquoi l'expérience visuelle dans ce contexte précis de perception des couleurs est illusoire. On voit mal cependant sur quelle base rationnelle il me serait possible de remettre en question le jugement (5) : si je considère que l'énonciateur est sincère, et que son jugement se fonde vraiment sur l'introspection, il semble impossible de le critiquer, et je me dois de lui accorder le dernier mot sur la question (Bar-On 2004).

De nouveau, Descartes est le premier à avoir mis en évidence l'importance de la perspective en première personne dans la connaissance introspective. Les différentes cogitationes que Descartes énumère au début de la Seconde méditation, et qu'il considère comme immunisées vis-à-vis du doute, sont formulées à l'aide de la première personne. Comme plusieurs commentateurs l'ont souligné, l'usage de la première personne est essentiel pour que l'argument de Descartes fonctionne (Hatfield, 2002 ; Williams, 1978). De nouveau, cela distingue la connaissance introspective d'autres formes de connaissance de soi. Supposons par exemple que René Descartes enquête sur sa vie passée, afin d'écrire son auto-biographie, et qu'il retombe par hasard sur un brouillon consignant les raisons de prudence qui l'ont conduit à ne pas publier son traité du Monde. A partir des données tirées de ce vieux document, Descartes peut conclure aussi bien (6) que (7).

(6) René Descartes a décidé par prudence de ne pas publier Le Monde.

(7) J'ai décidé par prudence de ne pas publier Le Monde.

Il s'agit dans les deux cas de l'expression d'une connaissance portant sur lui-même, la connaissance d'une décision qu'il a prise dans le passé et des motifs de cette décision, mais l'usage de la première personne n'est pas indispensable. En revanche, supposons maintenant que Descartes ait soudain le souvenir épisodique conscient d'avoir pris cette décision, le souvenir d'avoir craint d'être inquiété par l'Inquisition s'il publiait cet ouvrage. L'occurrence consciente de ce souvenir semble le justifier à juger :

(8) Je me souviens d'avoir décidé par prudence de ne pas publier Le Monde.

Ici cependant, le pronom de première personne joue un rôle essentiel. Le souvenir épisodique ne pourrait en effet pas justifier directement (9) :

(9) René Descartes se souvient d'avoir décidé par prudence de ne pas publier Le Monde.

Il semble en effet qu'un tel jugement ne puisse être justifié à partir du souvenir que si Descartes rajoute la prémisse :

(10) Je suis René Descartes.

Il en va exactement de même du Cogito : un épisode conscient de doute justifie le jugement en première personne « Je doute », mais pas les jugements en deuxième ou troisième personne ayant les mêmes conditions de vérité.

1. La notion d'accès privilégié

Descartes associe donc trois caractéristiques à la connaissance introspective : elle est immédiate, d'un grand degré de sûreté, et subjective, en ce sens qu'elle est associée à des croyances en première personne. On ne trouve néanmoins pas dans son œuvre de théorie de l'introspection, et il n'utilise d'ailleurs pas cette expression, quoiqu'il parle de la « conscience ou (…) témoignage intérieur que chaque homme trouve en lui-même quand il examine une observation quelconque » (La recherche de la vérité, AT 10, 524). Pourtant, une certaine image de l'introspection dite « cartésienne » s'est imposée dans la littérature contemporaine,

en particulier sous l'influence de Gilbert Ryle qui en fait sa cible principale dans (Ryle 2013/1949). Le texte dans lequel Ryle caractérise l'introspection au sens « cartésien » mérite d'être intégralement cité :

« Une personne est généralement censée être capable d'exercer de temps à autre une sorte spéciale de perception, nommée perception interne, ou introspection. Elle peut regarder ce qui se passe dans son esprit, quoique ce ne soit pas en un sens optique. Elle peut non seulement inspecter visuellement une fleur grâce au sens de la vue, écouter et discriminer les notes produites par une cloche par l'intermédiaire de l'ouïe ; mais elle peut aussi observer, de façon réflexive ou introspective, et sans l'intermédiaire de quelque organe sensoriel corporel que ce soit, les épisodes actuels de sa vie intérieure. Cette auto-observation (self-observation) est aussi communément supposée être immunisée à l'illusion, à la confusion, ou au doute. (…) Les perceptions sensorielles sont susceptibles d'erreur ou de confusion, mais ce n'est pas le cas de la conscience et de l'introspection » (Ryle 2013/1949, 14).

Ce texte ne reprend pas seulement les trois caractéristiques énumérées plus haut — immédiateté, sûreté, subjectivité — mais il prétend les expliquer à l'aide d'un modèle de l'introspection, le modèle de l'accès privilégié à un monde intérieur. Selon Ryle, le sujet cartésien peut « regarder ce qui se passe dans son esprit », et cela le met en relation avec des données auxquelles les autres agents n'ont pas accès. La reconstruction par Ryle du concept cartésien de l'introspection repose ainsi sur l'idée selon laquelle l'introspection est une perception d'objets internes, mais une perception d'un genre spécial, puisque « la conscience » et « l'introspection » ne sont susceptibles ni « d'erreurs », ni de « confusion ». Les données introspectives sont donc, d'après cette lecture, incorrigibles : personne d'autre que le sujet n'a accès à des raisons permettant de les mettre en question. Ce caractère incorrigible de l'introspection est censé expliquer l'autorité particulière que les sujets possèdent lorsqu'ils s'auto-attribuent des états conscients de manière introspective.

1. L'espace des positions sur l'introspection

Le premier problème que les théories philosophiques de l'introspection doivent examiner consiste à répondre à une question d'existence : la connaissance introspective, telle que nous venons de la caractériser — donc une connaissance immédiate, particulièrement sûre, qui confère un privilège et une autorité particulière au sujet — existe-t-elle vraiment ? N'est-ce pas un mythe philosophique ? Après tout, le concept d'introspection est un concept technique, qui a pris un sens très spécifique dans l'histoire de la philosophie après Descartes, et il est tout à fait possible qu'aucune connaissance ne corresponde réellement à ce concept. Les approches que nous qualifierons de « déflationnistes » (Wright 1998 ; Bar-On 2004) soutiennent que les phénomènes que l'on regroupe sous l'appellation d'introspection ne relèvent pas de la connaissance, mais plutôt d'une forme de comportement.

Les approches « réalistes » ont toutes pour point commun de considérer qu'il y a bel et bien des connaissances introspectives. Elles diffèrent cependant profondément entre elles dans la manière dont elles caractérisent ces connaissances. Le premier point de désaccord, qui est aussi le plus important, concerne le statut inférentiel ou non des connaissances introspectives. Nous connaissons les états d'esprit de nos semblables par inférence, en partant de

l'observation de leurs comportements. Or, selon l'image cartésienne de l'introspection, celle-ci ne repose précisément pas sur des inférences, et c'est d'ailleurs ce qui explique son immédiateté et sa grande sûreté. Nous débuterons notre discussion des approches réalistes par une présentation des théories non-inférentialistes, qui reprennent toutes à leur compte au moins une partie de l'héritage cartésien, et qui sont aujourd'hui les plus populaires parmi les philosophes. A l'intérieur de ce courant, il faut cependant introduire une distinction supplémentaire entre les auteurs qui adhèrent à l'image de l'introspection comme accès à un monde intérieur, que nous regrouperont sous l'appellation d'« introspectionnistes », et les auteurs qui rejettent cette image, que nous appellerons les « rationalistes ». Les auteurs introspectionnistes sont les plus proches de Descartes : ils acceptent non seulement l'idée que l'introspection est une façon non-inférentielle d'acquérir des connaissances sur ses états mentaux, mais aussi en un certain sens l'idée d'un accès privilégié du sujet à ses propres états d'esprit. Les rationalistes ne sont pas tous anti-cartésiens — Christopher Peacocke qui est un des principaux représentant de ce courant a même publié une défense du Cogito (Peacocke 2012, 2014) —, mais ils rejettent l'idée d'un accès privilégié à un monde intérieur. Pour eux, la connaissance de soi introspective ne présuppose pas que nous détournions notre attention du monde extérieur. Selon Gareth Evans (Evans 1982), dont l'œuvre est la source principale du courant rationaliste, l'esprit est transparent, car nous portons notre attention, lorsque nous formons une croyance introspective, exactement sur les mêmes phénomènes que ceux sur lesquels portent nos croyances non-introspectives. Peut-on parler d'« introspection » si l'on rejette l'idée d'un accès privilégié à l'intériorité ? Pas si l'on entend par là, à la façon des introspectionnistes, une attention portée à son propre esprit. Néanmoins, les rationalistes considèrent qu'il existe une connaissance qui possède les caractéristiques principales associées à l'image cartésienne : elle est immédiate et pas inférentielle, particulièrement sûre, et subjective. Et en ce sens, il est incontestable qu'ils prolongent l'héritage cartésien.

Enfin la dernière position, l'inférentialisme, est sceptique dans son esprit, et soutient qu'il existe bien des connaissances introspectives, mais que l'introspection n'a pas forcément toutes les propriétés que nous venons de décrire, et qu'elle ne constitue sans doute pas une classe naturelle, possédant une unité. Selon cette dernière approche, qui était celle de Gilbert Ryle (Ryle, 2013/1949) et qui a connu un puissant regain d'intérêt à la fois en psychologie et en philosophie (Carruthers, 2011 ; Cassam, 2014), il n'y a pas de différence de nature entre les connaissances que nous formons sur nos propres états mentaux et celles que nous formons sur les états d'autrui. Voici donc comment l'on peut résumer l'espace des différentes positions possibles sur l'introspection :

3. **Les approches déflationnistes**

Nous allons commencer par discuter le problème philosophique préalable portant sur l'existence même d'une connaissance introspective. Nous parlons de « connaissance introspective » à propos des vérités exprimées par des énoncés auto-attributifs du type suivant, qui décrivent en première personne ce que Descartes appelait les cogitationes dans la Seconde Méditation :

(11) Je ressens une intense douleur dans le bas du dos.

(12) Je vois plus nettement le pic vert qui se trouve sur la gauche que le geai sur la droite.

(13) Je sens la colère monter en moi.

(14) J'ai l'intention de faire un esclandre lors de notre prochaine réunion.

(15) Je crois qu'une troisième guerre mondiale aura lieu bientôt.

Dans la tradition ouverte par les derniers travaux de Wittgenstein, ces énoncés sont appelés des « aveux » — un terme qui traduit, l'anglais « avowal », lui-même introduit d'abord par Gilbert Ryle, et qui est aussi la traduction des expressions allemandes « Äußerung » ou « Ausdruck » utilisées par Ludwig Wittgenstein. De nombreuses sortes d'états ou d'événements mentaux peuvent être exprimés par les énoncés de ce type : des attitudes propositionnelles, c'est-à-dire des états décrits en relation à des contenus propositionnels, comme les croyances ou les intentions, des émotions, des sentiments, des sensations, ou même des épisodes mentaux ou des actions mentales. Ils manifestent l'autorité de la première personne sur le contenu de ses propres états mentaux : il serait en effet incongru de remettre en question la parole d'un agent considéré comme sincère et s'exprimant à l'aide d'un tel énoncé. Il existe à cet égard un contraste important entre les aveux d'une part, et les énoncés décrivant des états corporels de l'autre. Supposons par exemple que je ressente une douleur au pied, et que j'en déduise que mon pied est blessé. Suite à un examen, un médecin peut me corriger : il peut me demander quelles raisons fondent ma croyance en l'existence d'une blessure pour éventuellement critiquer ces raisons. En revanche, nul ne s'attendrait à ce qu'un médecin remette en question un aveu comme (11), exprimant l'existence non pas d'une blessure, mais d'une sensation de douleur. Le jeu de langage dans lequel s'insèrent les aveux ne semble faire aucune place à l'examen des raisons fondant, ou justifiant, ces énoncés, et il y a, à cet égard, une dissymétrie très claire entre les aveux et d'autres énoncés décrivant les états mentaux (Bar-On 2004 ; Falvey 2000 ; Wright 1998). Dans certains contextes en effet, les auto-attributions d'états mentaux sont tout aussi susceptibles d'être fondés sur des données probantes que les descriptions d'états corporels. Supposons ainsi qu'en regardant une vieille vidéo de campagne électorale, une femme politique affirme :

(16) J'avais visiblement très envie d'être élue députée.

Il s'agit d'une auto-attribution, mais d'une auto-attribution qui peut se fonder sur des données — ici sur l'observation par cette personne de son comportement tel qu'il est présenté par la vidéo. Ces raisons pourraient être discutées, comme n'importe quelles raisons fondant une affirmation particulière. Mais cela ne semble pas le cas si le même énoncé est utilisé au présent :

(17) J'ai très envie d'être députée.

Il est bien sûr possible de répondre à (17) en posant une question du type « Pourquoi ? », mais la question ne pourra pas être comprise comme demandant des raisons de l'auto-attribution : la femme politique y répondra en donnant les raisons qu'elle a d'avoir envie d'être députée, non en donnant les raisons qu'elle a de croire qu'elle a envie d'être députée. Notons qu'en revanche, la question « pourquoi ? » formulée en réponse à (16) ouvre la porte aux deux sortes de réponses : elle peut répondre en justifiant son envie d'être députée à l'époque, mais aussi en justifiant sa croyance qu'elle avait envie d'être députée à l'époque.

On retrouve ici un contraste souligné par Wittgenstein entre « l'usage comme objet » du pronom « je », et son « usage comme sujet » (Wittgenstein 1972) : dans les énoncés (11) à (15), le pronom est utilisé comme sujet car, pour reprendre l'expression de Wittgenstein, on n'a pas « procuré de possibilité d'erreur ». En revanche une possibilité d'erreur existe bien dans (16), puisque la locutrice pourrait s'être incorrectement identifiée sur la vidéo.

Enfin, les aveux sont particulièrement sûrs. Comparons par exemple (11) à l'énoncé (18), formulé sur la base d'une sensation corporelle :

(18) J'ai les jambes croisées.

Il s'agit d'un énoncé très sûr, mais il est néanmoins possible de corriger l'énonciateur dans certains contextes spécifiques, par exemple dans des situations anormales de perception des membres. En revanche, il est difficile de concevoir une situation où (11) pourrait être remis en question. Comme Crispin Wright le souligne (Wright, 1998), les aveux ne laissent aucune place au doute ou à l'indécision. Dans certains contextes, je peux me demander si j'ai ou non les jambes croisées ; en revanche, je ne peux pas me demander si je ressens ou non une douleur intense dans le bas du dos, ni si je vois ou non le pic vert qui picore dans mon jardin.

Les tenants de l'approche dite « expressiviste » de la connaissance de soi (Bar-On 2004; Wright 1998; Finkelstein 2003, 2010) partent d'une opposition tranchée avec le modèle cartésien de l'accès privilégié au monde intérieur (cf. Section 2). Selon ce modèle cartésien, les caractéristiques spécifiques des aveux proviennent de la perspective privilégiée qu'a le sujet pour observer ses propres états mentaux. (Wright 1998) illustre l'idée cartésienne d'une position d'observation privilégiée par l'image d'un kaléidoscope que seul le sujet serait en position d'observer. Dans un tel contexte, l'autorité de la première personne provient simplement du fait que le sujet est le seul à pouvoir observer une certaine scène visuelle. Par ailleurs, en l'absence d'informations visuelles sur la scène perçue dans le kaléidoscope, les interlocuteurs du sujet ne sont pas en position de remettre en question son témoignage. Bien entendu, souligne Wright, l'image a ses limites : alors qu'un kaléidoscope peut changer de mains, seul le sujet a accès de manière privilégiée à ses propres états, à son monde intérieur pour ainsi dire. Mais c'est précisément là que le modèle cartésien de l'accès privilégié se heurte à une difficulté : il doit ou bien postuler une forme entièrement spécifique de relation épistémique, capable de fonder les connaissances introspectives que le sujet peut former sur son monde intérieur (cf. Section 4.1) ; ou alors il doit renoncer à considérer la dissymétrie entre les aveux et les attributions en troisième personne comme vraiment radicale (cf. Section 4.2). Au fond, si l'autorité des sujets ne provient que de leur perspective privilégiée sur leur vie psychologique, cette autorité n'est que relative, et non absolue, et les spécificités des aveux ne sont pas expliquées de manière satisfaisante. Or les expressivistes, à tort ou à raison (cf. Section 6), entendent maintenir la thèse de l'autorité spécifique d'un agent sur ses états mentaux.

L'interprétation déflationniste des aveux entend renverser complètement le modèle cartésien. L'idée centrale de cette interprétation est la suivante : les aveux n'expriment tout simplement pas la connaissance de soi du sujet. On peut préciser cette idée centrale ainsi :

(Déflationnisme) Un aveu du type « Je E que P », où « E » décrit un état mental, et « P » le contenu de cet état, ne constitue pas une assertion selon laquelle le sujet se trouve dans l'état mental décrit, mais plutôt la simple expression de cet état mental.

L'interprétation déflationniste s'inspire fortement du paragraphe 244 des Recherches philosophiques de Wittgenstein. Voici ce qu'écrit ce dernier à propos de l'expression de la douleur :

« Une possibilité est que les mots soient reliés à l'expression originelle, naturelle, de la sensation, et qu'ils la remplacent. Un enfant s'est blessé, il crie ; et alors les adultes lui parlent, ils lui apprennent des exclamations, et plus tard des phrases. Ils enseignent à l'enfant un nouveau comportement de douleur. ' Tu dis donc que le mot "douleur" signifie en réalité crier ? ' — Je dis au contraire que l'expression verbale de la douleur remplace le cri et qu'elle ne le décrit pas ». (Wittgenstein 1953)

Un aveu comme « je ressens de la douleur », selon Wittgenstein, ne constitue pas une assertion décrivant un état mental, mais plutôt une expression qui, dans le cadre constitué par un certain jeu de langage, « remplace le cri ». Dans sa version la plus radicale, dite « expressivisme simple » (Finkelstein 2010), cette interprétation des aveux conduit à nier qu'ils expriment des propositions et donc qu'ils aient des conditions de vérité — une thèse qui rappelle l'expressivisme moral de A. J. Ayer (Ayer 2012), selon lequel les énoncés moraux ne décrivent pas des faits, mais manifestent plutôt nos réactions émotionnelles d'approbation ou de désapprobation, vis-à-vis des actions. La position expressiviste présente l'avantage d'expliquer très simplement pourquoi les aveux ne peuvent être remis en question : cela n'a tout simplement pas de sens de remettre en question par des raisons l'expression d'un état mental, pas plus que l'on ne peut répondre par des raisons à l'expression comportementale d'une émotion par exemple. Dans cette version radicale, elle est cependant peu plausible, pour des raisons exactement identiques à celles formulées par Peter T. Geach contre l'expressivisme moral (Geach 1965). Il est possible de construire des constructions grammaticalement complexes en combinant les aveux à des temps linguistiques, des opérateurs modaux, ou des connecteurs propositionnels. Ainsi :

(19) Dans deux jours, je ressentirai encore de la douleur.

(20) Mon médecin sait que je ressens de la douleur.

(21) Si je ressens de la douleur, je dois prendre mon médicament.

Il est impossible d'interpréter les énoncés (19) à (21) comme de simples expressions : il s'agit bien d'affirmations dotées de conditions de vérité. Or, si l'on accepte que ces énoncés possèdent des conditions de vérité, on voit mal comment on pourrait refuser d'attribuer également des conditions de vérité aux « aveux » qui y figurent comme constituants syntaxiques. Si les aveux ne possédaient pas de conditions de vérité, ils ne pourraient pas, par exemple, faire l'objet d'une négation. Ainsi, si l'on rejette l'aveu d'un jeune enfant en lui disant : « non, tu ne ressens pas vraiment de douleur », il semble bien que l'on nie son assertion, et non que l'on se situe dans un jeu de langage uniquement fondé sur l'expression.

Il existe cependant une version plus élaborée du déflationnisme, qui repose sur la distinction que l'on peut opérer entre un acte d'énonciation et le résultat de cette énonciation. Ainsi (Bar-On 2004) soutient-elle qu'on peut maintenir l'interprétation déflationniste à propos des actes intentionnels au travers lesquels les aveux sont produits, tout en admettant que les énonciations concrètes résultant de ces actes sont bien évaluables quant à la vérité. Cette position n'est pas absurde, puisqu'on sait bien qu'il faut distinguer entre le sens littéral d'un énoncé et l'information communiquée par une utilisation concrète, dans un contexte communicationnel précis, de l'énoncé en question. Par exemple, l'énoncé « Peux-tu fermer la fenêtre ? » sert dans la plupart des contextes à donner un ordre, quoiqu'il ait le sens littéral d'une question. Par ailleurs, le « néo-expressivisme » de Bar-On prétend expliquer certaines des caractéristiques des aveux sans présupposer l'existence d'un point de vue épistémique privilégié du sujet. Si un aveu découle directement de l'état mental qui l'exprime, on doit pouvoir expliquer selon elle pourquoi nous présupposons que les aveux ne doivent jamais être remis en question, et ceci sans céder à l'image cartésienne.

Le néo-expressivisme est bien plus plausible que l'expressivisme radical. Néanmoins, la subtilité et la modestie de cette position se retournent d'une certaine manière contre elle. Les néo-expressivistes admettent que les aveux ont des conditions de vérité, et donc qu'ils décrivent bien des faits psychologiques. Ils rejettent simplement la thèse selon laquelle leurs spécificités proviendraient de l'existence d'une méthode d'acquisition de connaissance privilégiée, introspective, portant sur ces états mentaux. On peut, selon eux, à la fois reconnaître qu'une connaissance des états mentaux est possible, et défendre une interprétation déflationniste des aveux selon laquelle ceux-ci n'ont pas en général pour rôle communicationnel d'exprimer une telle connaissance. Aucun éclaircissement n'est apporté sur la connaissance de soi introspective et sur sa nature spécifique. Il s'agit au fond plus d'une théorie des aveux et de leur usage dans la communication que de l'introspection. Cette conception peine en conséquent à expliquer les liens qui semblent exister entre certains aveux au moins et la production de connaissance. Si j'affirme, dans mon for intérieur, que je ressens de la douleur dans le bas du dos, je peux semble-t-il déduire de cette affirmation qu'au moins une personne dans le monde souffre de douleur dans le bas du dos. Mais comment est-ce possible, si l'on n'admet pas pour commencer que l'aveu lui aussi exprime une connaissance ? Il est assez douteux qu'on puisse réellement déconnecter les auto-attributions d'états mentaux de tout lien avec la production de connaissance.

Nous allons donc maintenant présenter les théories qui adoptent une interprétation réaliste des aveux, au sens où elles considèrent qu'il existe bel et bien des méthodes spécifiques d'acquisition de connaissances en première personne sur nos propres états mentaux. Nous commencerons par présenter les théories les plus proches de l'héritage cartésien, les théories « introspectionnistes », qui acceptent l'idée d'un accès privilégié du sujet à son monde intérieur.

4. **Les approches introspectionnistes**

On peut distinguer deux grandes approches introspectionnistes, dont l'une trouve sa source dans les œuvres de Descartes et Russell, et l'autre dans celles de Locke (Gertler 2011). Ces deux approches s'accordent sur l'idée qu'il existe bien une relation épistémique unique, privilégiée, entre une personne et ses propres états mentaux, qui explique son autorité particulière vis-à-vis de ceux-ci. Elles se différencient cependant sur un point central : selon la

première approche, l'introspection doit être conçue comme une relation directe avec les états mentaux, irréductible à quelque mécanisme perceptif que ce soit. Selon la seconde en revanche, l'introspection doit être conçue sur le modèle de la perception, et peut donc être étudiée dans un cadre naturaliste, voire mécaniste. Nous nommerons la première approche la théorie de l'introspection comme connaissance directe, et la seconde la théorie de l'introspection comme sens interne.

1. Les théories de l'introspection comme connaissance directe

Selon Descartes, il existe une différence de nature et non simplement de degré entre la connaissance qu'une personne peut former sur ses propres états mentaux, et les connaissances qu'elle peut former sur les objets extérieurs. Comme le montre la démarche du doute radical dans les Méditations métaphysiques, on peut remettre radicalement en question les connaissances portant sur le monde, y compris les connaissances issues des organes sensibles, puisqu'on peut confondre une expérience sensorielle non-véridique, vécue par exemple pendant un rêve, avec la perception d'une réalité. Nos organes sensoriels sont des mécanismes qui, sur la base d'interactions causales avec l'environnement, nous permettent de former des représentations mentales — des idées, dans le vocabulaire de Descartes. Or, un mécanisme fondé sur la causalité peut mal fonctionner, comme le montrent les situations d'illusion ou d'hallucination. En revanche, la connaissance que j'ai, en tant que sujet et en première personne, de ma pensée au moment où je pense, ne laisse pas de place au doute. C'est la raison pour laquelle Descartes soutient que l'esprit, comme nous l'avons vu plus haut, « est plus aisé à connaître que le corps ». Selon lui, l'introspection fonde en fait toute la connaissance humaine. Dans la perspective de la seconde Méditation, on ne peut parler de connaissance qu'à condition que le sujet connaissant ait un accès introspectif à des raisons de croire. Même une connaissance perceptive — disons par exemple la connaissance que j'acquiers par la vision à propos des propriétés sensibles d'un morceau de cire — n'est rationnellement justifiée qu'à partir du moment où je peux, en tant que sujet, réfléchir aux raisons que m'apportent les organes sensibles, et les mettre en relation avec d'autres raisons. En ce sens l'œuvre de Descartes est l'une des sources du courant internaliste en philosophie de la connaissance, puisqu'il considère que connaître suppose de pouvoir avoir un accès conscient à des raisons de croire. Mais ces raisons, pour valoir comme telles pour le sujet, doivent lui être accessibles par introspection.

Descartes présuppose dans son œuvre qu'il existe un accès privilégié du sujet à ses pensées, mais il n'explique pas la nature de cet accès : on ne trouve pas à proprement parler chez lui de théorie de l'accès introspectif. Bertrand Russell développe une telle théorie dans les Problèmes de philosophie (Russell 1989). L'idée centrale de cette théorie est la suivante : il existe une relation métaphysique spécifique et irréductible d'« acquaintance » entre les sujets et certains objets, dont les états mentaux. C'est l'existence de cette relation qui explique que nous puissions connaître ces états directement, immédiatement, sans aucune inférence, et avec une autorité caractéristique. Selon Russell, « nous avons l'expérience directe (acquaintance) d'une chose quand elle est là directement devant nous, que nous en avons conscience, sans l'intermédiaire d'aucun processus d'inférence ou de quelque connaissance de vérité que ce soit » (Russell 1989, tr. fr. 69). Parmi les choses dont nous avons une connaissance directe figurent d'abord les sense-data, c'est-à-dire les propriétés sensibles associées à des expériences conscientes comme l'expérience d'une nuance de couleur, ou de l'intensité d'un son. Russell considère que les sense-data ne sont pas des entités mentales —

ou du moins, qu'ils ne sont pas connus par les sujets comme des entités mentales. Il affirme cependant que nous connaissons aussi nos propres états mentaux par acquaintance :

« Nous n'avons pas seulement conscience des choses, mais nous avons souvent conscience d'avoir conscience d'elles. Quand je vois le soleil, j'ai souvent conscience de voir le soleil ; et ainsi 'le fait que je vois le soleil' est un objet dont j'ai l'expérience directe. » (Russell 1989, tr. fr. 72)

Dans un autre texte, Russell souligne que nous pouvons faire directement l'expérience non seulement des états perceptifs, mais de tous les états mentaux. Il prend l'exemple du désir, et affirme que nous avons directement conscience, par acquaintance, de notre désir de nourriture lorsque nous avons faim. Il soutient qu'il existe une dissymétrie entre les connaissances que nous pouvons former sur les faits extérieurs d'un côté, y compris les faits portant sur les états mentaux de nos semblables, et les auto-attributions reposant sur l'acquaintance. Les secondes, parce qu'elles reposent sur une connaissance directe, ne laissent aucune place au doute : selon Russell, la possibilité du doute est la marque qu'une connaissance est descriptive et non directe (Russell 1989). Je ne peux donc pas savoir que j'ai mal par acquaintance, et dans le même temps douter de l'existence de ma douleur. En revanche, nos connaissances des objets extérieurs peuvent toujours être remises en question, parce qu'elles ne sont pas directes mais reposent sur des inférences. Russell prend l'exemple d'une table en train d'être perçue visuellement par un sujet : alors qu'on ne peut remettre en question l'existence ni des sense-data constituant l'expérience de la table, ni de l'acte conscient de voir, on peut douter que la table réelle existe. Elle est, dit Russell, « le résultat d'une inférence à partir de ce qui est immédiatement connu » (Russell 1989, trad. fr. 33).

Les théories contemporaines de l'acquaintance partagent l'idée centrale de Russell, selon laquelle tout sujet a un accès privilégié à ses états mentaux en vertu de l'existence d'une relation irréductible susceptible de lui procurer des justifications non-inférentielles de ses croyances introspectives. En général, elles évitent cependant de postuler l'existence de sense-data. La thèse selon laquelle les relata de la relation d'acquaintance sont des sense-data, des objets privés au statut ontologique problématique, n'est en fait pas essentielle à l'épistémologie de la connaissance directe. Ainsi les auteurs qui se rattachent actuellement à la tradition russellienne (Chalmers 2003 ; Conee 1994 ; Balog 2012 ; Fumerton 1995 ; Gertler 2001) considèrent que nous connaissons des états sensoriels et non des sense-data dans la connaissance introspective fondée sur l'acquaintance. Ils soulignent souvent, en s'appuyant sur (Kripke 1980) qu'il n'existe pas de distinction entre l'apparence d'une sensation et sa réalité : une sensation, lorsqu'elle est ressentie comme douloureuse, est ipso facto douloureuse. Cela n'est pas supposé montrer que l'introspection est infaillible, mais plutôt que cette sorte de connaissance ne repose pas sur des données observables, et qu'elle est donc bien directe.

Ces approches se heurtent à plusieurs difficultés. En premier lieu, la relation d'acquaintance est supposée métaphysiquement primitive. Elle est conçue comme irréductible, et donc comme inexplicable par quelque mécanisme causal que ce soit. On peut soupçonner que l'acquaintance n'est qu'un nom qui cache notre ignorance des mécanismes qui rendent possible la connaissance introspective. Afin d'expliquer l'existence d'une sorte de connaissance problématique, on stipule qu'il existe une relation épistémique directe qui

fonde cette sorte de connaissance. Mais a-t-on expliqué quoi que ce soit par cette stipulation ? Les philosophes naturalistes, parce qu'ils adoptent le cadre explicatif des sciences cognitives, sont particulièrement sensibles à cette objection : il semble très difficile de concilier une conception physicaliste de la connaissance, compatible avec les connaissances psychologiques et neuroscientifiques contemporaines, avec l'idée même d'acquaintance (Finkelstein 2003). En second lieu, les théories reposant sur l'acquaintance sont solidaires, depuis Descartes, d'une conception internaliste de la connaissance. Une connaissance introspective, dans ce cadre, est une croyance portant sur un état mental du sujet, fondée sur une raison qui lui soit accessible. Cela soulève de sérieuses difficultés. Si, pour être des connaissances, nos croyances introspectives doivent être justifiées par des raisons elles-mêmes accessibles par introspection, n'y a-t-il pas là une forme de circularité vicieuse ? Mais surtout, quelles sont ces mystérieuses raisons supposées fonder les connaissances introspectives ? Les « aveux », ces énoncés qui expriment notre connaissance introspective, semblent sans fondement, au sens où il paraît déraisonnable de demander qu'ils soient justifiés (cf. Section 3). Pourquoi est-ce le cas, cependant, si l'acquaintance fournit des raisons ? Enfin, connaître suppose d'abord de conceptualiser (Gertler 2011). Mais d'où proviennent nos concepts de sensation, si ce n'est d'une forme inexpliquée d'observation interne ? Russell lui-même reprend à son compte l'expression de « sens interne » pour désigner la méthode d'acquisition de connaissance fondée sur l'acquaintance. Il est pourtant paradoxal, sinon contradictoire, de comparer l'acquaintance à un « sens », puisque celle-ci est supposée ne pas reposer sur des interactions causales avec l'environnement. Les tenants de la théorie de l'acquaintance doivent donc expliquer comment l'esprit peut parvenir à former des concepts des sensations sans interagir causalement avec elles, ce qui constitue un défi important pour cette approche. La problématique de l'introspection rejoint ici celle des concepts dits « phénoménaux » (Chalmers 2003 ; Gertler 2001, 2011).

1. Les théories du sens interne

La différence fondamentale entre les théories du sens interne et les théories de l'introspection comme connaissance directe tient en une idée, qui remonte à l'œuvre de Locke : l'analogie entre l'introspection et la connaissance sensorielle. Dans un texte célèbre de l'Essai concernant l'entendement humain, Locke soutient qu'il existe une source de connaissance « très semblable à un sens », qui donne un accès privilégié, en première personne, au domaine psychologique :

« L'autre source, dans l'expérience, dont l'entendement tire ses idées réside dans la perception, en nous, des opérations de notre propre esprit. De là proviennent des idées que nous ne pourrions dériver des choses externes — telles que les idées de la perception, de la pensée, du doute, de la croyance, de la connaissance, du vouloir, et de toutes les choses diverses que fait l'esprit. En prenant conscience de ces actions de notre esprit, et en les observant en nous, notre entendement en tire des idées qui sont aussi distinctes que celles qu'il tire des corps qui affectent nos sens. Cette source d'idée se trouve dans le for intérieur de tout homme. Et quoiqu'elle ne puisse s'identifier à un sens, puisqu'elle n'a rien à voir avec les objets extérieurs, elle s'apparente fortement à un sens, et pourrait être proprement nommée 'sens interne'. » (Locke 2008, II, 1, 4).

Selon Locke, il existe, à côté des sensations, un autre mode de perception qu'il nomme « réflexion » et qui permet à un sujet d'acquérir des représentations des processus

psychologiques qui adviennent dans son esprit. La théorie lockéenne est complexe, et elle est traversée de tensions qui la rendent difficile à interpréter (Scharp 2008). Plusieurs philosophes contemporains reprennent cependant l'idée d'un sens interne, et ce sont ces conceptions modernes de l'introspection comme mécanisme de perception interne que nous allons discuter (Armstrong 2003 ; Goldman 2006 ; Lormand 1996 ; Lycan 1996; Nichols et Stich 2003). L'idée centrale de ces approches contemporaines peut être formulée ainsi : il existe, au fondement de la connaissance de soi introspective, une forme de conscience de soi qui n'est rien d'autre qu'une perception par l'esprit de ses propres états (Churchland 1988). Or, on considère aujourd'hui que la perception peut être étudiée dans le cadre naturaliste des sciences cognitives. On peut donc reprendre l'intuition centrale des théories de l'acquaintance, selon laquelle la connaissance introspective est non-inférentielle, mais dans un cadre naturaliste, en considérant que cette connaissance est le produit d'un mécanisme sensoriel ad hoc, capable de produire des méta-représentations de nos représentations mentales en interagissant causalement avec elles. Ajoutons que ces approches adoptent en général une épistémologie externaliste, qui met l'accent sur la fiabilité du processus causal permettant de produire les croyances introspectives, plutôt que sur les raisons accessibles en première personne fondant ces croyances.

Les conceptions de l'introspection comme sens interne sont associées à une certaine conception de la conscience phénoménale. Par « conscience phénoménale », nous entendrons la propriété de certains états mentaux supposée expliquer que ces états aient un aspect subjectif, c'est-à-dire qu'ils soient associés à un certain effet subjectif que nous nommerons sa « phénoménologie ». Ainsi, un état de douleur est un état conscient au sens phénoménal, car cela fait un certain effet subjectif de vivre l'expérience correspondante : ressentir une douleur intense ou un plaisir intense, cela ne fait pas le même effet subjectif, car les phénoménologies associées sont distinctes. Quel lien peut-on faire entre la conscience phénoménale et la théorie du sens interne ? Pour répondre à cette question, distinguons deux sens du concept de conscience : ce concept peut décrire d'une part une caractéristique d'un état mental — le concept de conscience phénoménale appartient à cet usage —, mais aussi une caractéristique d'une personne. On peut dire en effet qu'une personne est consciente d'une certaine information, ou qu'elle prend conscience d'être dans un état psychologique donné, par exemple qu'elle prend conscience de la colère qui monte en elle. Or certains philosophes considèrent qu'il est possible d'expliquer la conscience d'état, la conscience phénoménale, à partir de la conscience comme « prise de conscience » par une personne. Voici une formulation précise de cette idée, que nous nommerons la « théorie du sens interne » (TSI) et que nous reprenons, en substance, à (Carruthers 2007) :

(TSI) Un état mental conscient au sens phénoménal est un état sensoriel, qui est (méta-)représenté par un état d'ordre supérieur lui aussi sensoriel, par un mécanisme de détection, le « sens interne ».

En anglais, deux expressions peuvent être utilisées pour traduire le mot français « conscience » : « consciousness » et « awareness ». Selon la théorie du sens interne, la conscience phénoménale s'explique par le fait que le sujet est réflexivement conscient, au sens de l'expression anglaise « awareness », de certains de ses états sensoriels. Ainsi David Armstrong écrit-il, dans l'une des premières défenses contemporaines de cette théorie « que la conscience (consciousness) n'est rien d'autre que la perception (awareness) des états

mentaux internes par la personne dont ce sont des états » (Armstrong 2003/1968, 94). Il ne suffit pas, autrement dit, qu'une personne ait des sensations pour que celles-ci soient conscientes au sens phénoménal : encore faut-il que le sujet en prenne conscience, et cela suppose l'exercice de ce « scanner » intérieur, de cette capacité à détecter les états occurrents, qu'est le sens interne. Dans cette perspective, c'est l'existence de ce mécanisme d'introspection qu'est le sens interne qui explique que certains de nos états mentaux soient conscients (Lycan 1996, 13). Il en découle immédiatement que lorsqu'un état est conscient, il peut faire l'objet de croyances de second ordre introspectives. Ainsi, nos croyances subsistent la plupart du temps sans que nous en prenions conscience, mais lorsque nous prenons conscience du contenu d'une croyance, lorsque celle-ci devient consciente, c'est que le sens interne la détecte, et des méta-représentations introspectives en découlent normalement (Nichols et Stich 2003). Nous allons maintenant discuter les difficultés auxquelles se heurte cette approche.

La première difficulté concerne la phénoménologie (Dretske 1995 ; Güzeldere 1995). On considère traditionnellement que les sens externes se distinguent les uns des autres en vertu de ce qu'on nomme des « sensibles propres », c'est-à-dire des propriétés perceptibles associées de façon unique aux différentes modalités. Ainsi, la vue et elle seule permet de nous représenter les couleurs ; l'audition, les sons ; l'odorat, les odeurs ; le toucher, la dureté. Ces sensibles propres déterminent la phénoménologie spécifique d'une modalité sensorielle, dont on peut considérer qu'elle est absente de l'expérience consciente lorsque la modalité ne fonctionne plus. Parce que ses états sensoriels conscients ne peuvent représenter les couleurs, la phénoménologie associée à l'expérience consciente d'une personne non-voyante possède ainsi une spécificité, précisément parce que cette personne n'a pas accès à la phénoménologie propre des expériences de couleur. Mais si le sens interne est littéralement compris comme une modalité sensorielle, ne devrait-il pas y avoir également une phénoménologie spécifique associée à son fonctionnement ? Le sens interne, autrement dit, ne devrait-il pas nous présenter les états internes de nos esprits comme dotés de propriétés sensibles propres, de la même manière que la vision nous présente les surfaces et les volumes comme colorés ? Cette objection est parfois associée à la thèse de la transparence de l'expérience consciente (cf. Section 5). Si l'introspection supposait l'intervention d'un sens interne, ne devrait-on pas, pour former des connaissances réflexives sur ses états mentaux, détourner le regard des objets extérieurs et le diriger vers l'intériorité de l'esprit — quel sens que cette expression puisse avoir ? Or, comme le souligne, par exemple, Christopher Hill, « lorsqu'on essaie de porter son attention, dans l'introspection, sur une expérience perceptive (…) on est conscient uniquement de ce dont elle est une expérience » (Hill 2009). Il n'y a pas, pour reprendre l'expression de Gilbert Harman, de « peinture mentale » vers laquelle on pourrait tourner son attention et qui constituerait pour ainsi dire le sensible propre du sens interne (Harman 1990). Il est possible que cette objection repose sur un malentendu, voire sur une pétition de principe (Carruthers 2007). Par définition, pour les partisans de (TSI), un état conscient (conscious) est un état dont le sujet est conscient (aware). Il n'y a pas, autrement dit, de conscience de premier ordre sans conscience réflexive. Ce n'est donc pas comme si les états mentaux possédaient une phénoménologie, qui était ensuite modifiée par la méta-représentation produite par le sens interne : il n'y a tout simplement aucune phénoménologie consciente sans cette méta-représentation. Certes, il y a une importante différence de nature entre la perception par le sens interne et la perception des objets externes : le sens interne ne présente pas au sujet d'objets internes possédant des apparences

sensibles, et en ce sens précis il n'est pas associé à une phénoménologie (Lycan 1996). Pour autant, aucun état sensoriel ne serait associé à un effet subjectif sans l'intervention du sens interne. Comme Leibniz le remarque déjà dans les Nouveaux essais sur l'entendement humain, II, 1, §19 (Leibniz 1966), nous ne saurions réfléchir sur tous nos états mentaux sans que cela occasionne une régression à l'infini. Il s'en suit que les états issus des opérations du sens internes ne peuvent pas être eux-mêmes conscients. Mais cela n'est pas en soi problématique, au contraire : une force de (TSI) est de distinguer entre les états sensoriels, et les états sensoriels conscients. On s'accorde en effet aujourd'hui à considérer que tous les états sensoriels ne sont pas conscients. Ainsi, les patients dotés de vision aveugle, une condition causée par des lésions du cortex visuel, peuvent conserver de nombreuses capacités de discrimination visuelle, sans pour autant rapporter aucune phénoménologie. Par ailleurs, les travaux de psychologie cognitive sur l'amorçage montrent que des représentations sensorielles peuvent être activées dans le cerveau sans être conscientes pour autant (Dehaene et Naccache 2001 ; Naccache 2006). Le fait que (TSI) implique l'existence de représentations mentales inconscientes est donc en fait une force de cette hypothèse.

La seconde difficulté concerne un trait important de la connaissance introspective : sa très grande sécurité. Selon (TSI), l'idée selon laquelle nos auto-attributions introspectives seraient plus sûres que les jugements de perception est erronée, puisque le sens interne n'est rien d'autre qu'un mécanisme de perception des états de l'esprit par lui-même. Comme tout mécanisme perceptif, il peut mal fonctionner dans certaines circonstances, et dans ce cas les données acquises par introspection ne correspondront pas à la réalité du phénomène psychologique représenté. Autrement dit, il est possible selon (TSI) qu'il y ait un hiatus entre l'apparence d'un état mental introspecté et sa réalité, y compris dans le domaine sensoriel. Je peux donc prendre conscience que je ressens de la douleur sans réellement ressentir de douleur, ce qui revient à abandonner l'une des caratéristiques essentielles de l'image cartésienne de l'introspection. Cela ne signifie pas seulement que je peux ressentir de la douleur même lorsque mon corps n'a subi aucun dommage. L'existence d'expériences de douleur non-véridique, comme les douleurs subies par les patients amputés dans leur membre fantôme, est bien documentée et largement acceptée. Mais la conséquence de (TSI) est bien plus radicale : l'hypothèse implique qu'un sujet peut se tromper sur ses propres sensations au moment même où il a ces sensations. C'est ce point qui est problématique. Il semble difficile — voire impossible — de distinguer conceptuellement entre prendre conscience d'une douleur d'une part, et avoir mal de l'autre. Comment pourrait-on prendre conscience d'une sensation de douleur sans avoir mal ? Pourtant cette possibilité conceptuelle est une implication essentielle de (TSI), comme William Lycan le reconnaît d'ailleurs (Lycan 1996).

1. **Introspection et cécité à soi-même**

Dans le cadre de sa discussion très influente du modèle introspectionniste, Sydney Shoemaker a développé une critique particulièrement ambitieuse de l'idée même de sens interne. Son argument mérite une discussion séparée car il jette un éclairage intéressant sur les relations entre conscience, connaissance de soi, et introspection (Shoemaker 1996, 30). La cible de Shoemaker est la thèse de l'indépendance, que l'on peut formuler ainsi :

Thèse de l'indépendance : l'existence des états mentaux que nous connaissons de façon introspective est indépendante de nos croyances introspectives ainsi que des mécanismes qui permettent d'acquérir ces croyances (Shoemaker 1996, 224).

Selon les théories du sens interne, la thèse de l'indépendance est vraie. En effet selon ces théories les états qui font l'objet de connaissance introspective, et les états dont l'existence découle de cette connaissance introspective, sont des entités séparées. Supposons par exemple que je réalise soudain que j'ai mal au dos. Un théoricien du sens interne comme Lycan soutiendra que la douleur préexiste, en tant qu'état représentationnel, à sa détection par le sens interne. Lorsqu'elle est détectée la douleur devient consciente et elle peut donner lieu à des croyances introspectives, mais cela n'empêche pas qu'elle existait avant d'être détectée, sans que le sujet en ai conscience.

Afin de montrer que cette thèse est fausse, et donc que les théories du sens interne sont vouées à l'échec, Shoemaker introduit l'idée d'un agent « aveugle à lui-même » (self-blind). Un tel agent possède toutes nos capacités cognitives, tous nos concepts, toutes nos capacités de perception et de raisonnement, à l'exception du mécanisme de sens interne permettant de détecter les états mentaux par introspection. Les théories du sens interne impliquent que l'existence d'un tel agent est concevable, puisqu'elles considèrent l'introspection comme un mécanisme spécifique, qui pourrait donc être sélectivement supprimé. Ceci étant posé, Shoemaker soutient que l'hypothèse de l'existence d'un tel agent débouche sur une contradiction. Les considérations qu'il avance pour justifier cette thèse étant nombreuses et complexes, nous allons nous concentrer sur le cas spécifique des croyances, et analyser l'argument visant à établir que l'hypothèse d'un agent souffrant de cécité vis-à-vis de ses propres croyances est contradictoire.

Cet argument utilise le paradoxe de Moore. Considérons un énoncé de la forme paradoxale « p, mais je ne crois pas que p », par exemple :

(22) Il pleut, mais je ne crois pas qu'il pleuve.

Cet énoncé n'est pas contradictoire dans sa forme logique, puisque l'occurrence d'un événement météorologique comme la pluie est indépendante de l'occurrence de la croyance portant sur la pluie. Néanmoins on s'accorde en général à considérer qu'un locuteur rationnel ne pourrait pas être justifié à énoncer (22). Les raisons qui font de (22) un énoncé paradoxal sont complexes, mais pour les besoins de la discussion supposons à la suite de Shoemaker qu'un locuteur rationnel puisse a priori réaliser qu'il ne peut pas être justifié à produire (22). Or, soutient Shoemaker, un agent souffrant de cécité à soi devrait en théorie pouvoir formuler un énoncé tel que (22). Un tel agent possède d'une part des connaissances acquises en troisième personne, de façon non-introspectives, sur ses propres états mentaux ; et d'autre part bien entendu des connaissances portant sur l'environnement, par exemple portant sur le contexte météorologique. Selon Shoemaker il peut être justifié à soutenir qu'il pleut, mais néanmoins, étant atteint de cécité vis-à-vis de ses croyances, ne pas savoir qu'il croit qu'il pleut, et donc affirmer (22). Cette conséquence est cependant contradictoire avec l'hypothèse de départ : si l'agent aveugle à lui-même est parfaitement rationnel, un simple raisonnement a priori doit l'empêcher d 'accepter (22). En raison de cette contradiction, on peut déduire que l'hypothèse de la concevabilité de l'existence d'un agent aveugle à lui-même est fausse,

et donc que la thèse de l'indépendance est fausse également. De façon plus générale, Shoemaker conclut à propos de l'introspection que « la réalité connue et la faculté qui permet de la connaître sont pour ainsi dire faites l'une pour l'autre : aucune des deux ne pourrait exister sans l'autre » (Shoemaker 1996, 245), et qu'à cet égard l'introspection diffère de manière essentielle de la perception.

L'argument très ambitieux de Shoemaker présente cependant des failles qui l'invalident (Siewert 2003 ; Kind 2003 ; Peacocke 2008). En premier lieu, Shoemaker présuppose dans la partie de l'argument qui utilise le paradoxe de Moore que l'agent aveugle à lui-même devrait être susceptible d'accepter l'énoncé (22) comme étant justifié, ceci en vertu des déficiences épistémiques liées à la cécité à soi. Il est difficile, néanmoins, de voir comment il peut soutenir cette thèse. Certes, cet agent n'a pas d'accès introspectif, via le sens interne, à ses propres croyances. Mais cela n'implique pas qu'il n'ait pas d'accès à ses propres croyances par d'autres moyens que par le sens interne, par exemple par des inférences à partir de son propre comportement, de ses états corporels internes, ou de tout autre ensemble de données pertinentes. Shoemaker suggère qu'il est contradictoire qu'un agent aveugle vis-à-vis de lui-même puisse à la fois comprendre que l'énoncé de Moore est paradoxal, et n'avoir aucun accès à ses croyances par le sens interne, mais il est difficile de lui donner raison. Son raisonnement semble en effet reposer sur une confusion entre connaissance introspective et connaissance de soi : même si l'on suppose qu'il ne peut pas acquérir de connaissance par le sens interne, l'agent aveugle vis-à-vis de lui-même reste capable d'acquérir des connaissances sur lui-même par toutes sortes de méthodes (Siewert 2003, Kind 2003). Shoemaker semble penser qu'au moins dans le cadre du modèle introspectionniste, un agent aveugle à lui-même devrait nécessairement disposer de moins d'informations sur ses propres états mentaux qu'un agent capable d'utiliser le sens interne, et que cette déficience devrait pouvoir se manifester dans son comportement. Mais le « nécessairement » ne paraît guère justifié. Pour prendre une analogie, une personne non-voyante ne dispose pas nécessairement de moins d'informations sur les objets visibles et leurs propriétés visibles (forme, couleur, etc. …) qu'une personne capable d'utiliser normalement son système visuel, parce qu'il y a bien d'autre méthodes que la vision pour acquérir des informations sur le monde visuel (Kind 2003). Par ailleurs, la conclusion de Shoemaker paraît beaucoup trop forte (Siewert 2003 ; Peacocke 2008). Il semble en effet que chez beaucoup d'animaux, certains états mentaux qui dans l'espèce humaine sont accessibles par introspection peuvent exister sans introspection. Ainsi, il paraît plausible de soutenir qu'un chien soit susceptible d'éprouver de la douleur, mais il est moins clair qu'il soit capable de former des croyances réflexives sur ses expériences de douleurs. Cela va dans le sens de la thèse de l'indépendance, une thèse qui selon l'argument de Shoemaker devrait être non seulement fausse, mais nécessairement fausse.

Plusieurs auteurs ont proposé une lecture modeste de l'argument de Shoemaker, compatible avec la thèse de l'indépendance (Siewert 2003 ; Peacocke 2008). Christopher Peacocke soutient ainsi que l'argument de la cécité à soi ne montre pas qu'il existe un lien nécessaire entre les états de premier ordre et les croyances introspectives, mais plutôt un lien nécessaire entre les concepts de ces états et ces croyances (Peacocke 2008, 273). Il concède à Shoemaker qu'il est impossible qu'un agent possède à la fois toutes nos capacités cognitives et rationnelles, qu'il possède le concept d'un jugement conscient, et qu'il soit incapable de s'auto-attribuer ses jugements conscients — ce qui reste compatible avec la thèse de l'indépendance. Selon Peacocke, si je suis capable de juger consciemment qu'il pleut, et que

je possède le concept de croyance, alors nécessairement je dois aussi être capable de m'auto-attribuer ce jugement. Nous allons explorer en détail, dans la prochaine section, cette conception rationaliste de la connaissance introspective.

5. Les approches rationalistes

1. Evans et l'idée de transparence au monde

Les philosophes rationalistes partagent la méfiance des expressivistes vis-à-vis de l'image cartésienne d'un monde intérieur auquel le sujet aurait un accès privilégié en première personne, tout en maintenant qu'il peut y avoir des connaissances introspectives produisant des certitudes. On doit à Gareth Evans, dont l'œuvre est à bien des égards la principale source de ce courant de pensée, une formulation éloquente de cette méfiance (Evans 1982). Si l'on conçoit l'introspection sur le modèle d'un regard tourné vers le monde intérieur, alors, selon Evans, l'introspection n'existe tout simplement pas. Dans un passage souvent cité, il insiste sur le fait que notre connaissance réflexive se fonde toujours sur des raisons de premier ordre, portant sur le monde extérieur :

« Lorsque nous nous auto-attribuons une croyance, nos yeux sont, pour ainsi dire, et parfois même littéralement, tournés vers l'extérieur — vers le monde. Si quelqu'un me demande ' Pensez-vous qu'il y aura une troisième guerre mondiale ? ', je dois porter mon attention, afin de lui répondre, exactement sur les mêmes phénomènes extérieurs auxquels je porterais attention si je répondais à la question ' Y aura-t-il une troisième guerre mondiale ' ? » (Evans 1982, 225).

L'idée centrale d'Evans est en effet complètement opposée au modèle cartésien de l'accès privilégié : selon lui, il n'est pas nécessaire de prêter attention à d'hypothétiques objets mentaux pour acquérir des connaissances sur son propre esprit. L'état mental dont je prends connaissance ne devient jamais un objet pour moi, et a fortiori il n'est pas un objet de perception interne (Bermudez 2010). On peut donc parler, pour reprendre l'expression de (Bar-On 2004), d'une transparence au monde (transparency-to-the-world) de l'esprit. Cette conception de la transparence a l'avantage d'expliquer de manière simple la dissymétrie entre les aveux et les attributions d'états mentaux à autrui. Lorsque j'essaie d'identifier les états mentaux d'une autre personne, je le fais d'une manière indirecte, inférentielle, sur la base d'une observation de son comportement, et donc à partir de données. Je peux aussi prendre une telle perspective détachée sur mon propre comportement. Je peux par exemple prendre conscience, en réfléchissant à mes actions passées, que je me comporte de manière agressive vis-à-vis d'une certaine personne, et en déduire que j'éprouve une certaine aversion inconsciente à son égard. Néanmoins cette perspective détachée vis-à-vis de mes états mentaux n'est pas caractéristique de ma relation épistémique normale et habituelle à mes propres états, en tant précisément que je les considère comme mes états. Habituellement, je ne fais pas de différence, selon Evans, entre réfléchir à ce que je crois, et réfléchir à ce qui est vrai du monde extérieur : j'utilise, pour m'auto-attribuer une croyance réflexive, exactement les mêmes capacités cognitives que pour former une croyance de premier ordre, et je les dirige vers les mêmes données. L'autorité et l'immédiateté caractéristiques des auto-attributions s'expliquent donc par l'usage de la règle suivante : « chaque fois que l'on est en position d'affirmer que p, l'on est aussi ipso facto en position d'affirmer ' je crois que p ' » (Evans 1982, 225-226). D'où le sentiment que les auto-attributions sont infondées : elles n'ont

pas d'autres raisons, la plupart du temps, que celles qui fondent les jugements de premier ordre, et ne reposent donc pas sur de nouvelles données. D'où aussi, selon Evans, leur grande sûreté, puisqu'il suggère que la connaissance de soi acquise par cette procédure transparente est immunisée contre le doute. On voit donc que si Evans rejette l'idée d'accès privilégié à un monde intérieur, il maintient que les connaissances introspectives sont immédiates, particulièrement sûres, et qu'elles ont un lien essentiel à la première personne.

1. **Réflexion critique et aveux : la position rationaliste**

Deux questions se posent lorsqu'on prend la procédure d'Evans comme point de départ. En premier lieu, quel est le type de la connaissance acquise par cette procédure ? Evans insiste sur ce qu'elle n'est pas : il ne s'agit pas d'une connaissance reposant sur l'observation interne. Néanmoins, il n'apporte guère d'éclaircissements sur la nature de cette connaissance, ni sur les raisons pour lesquelles on peut bien la considérer comme une connaissance. En second lieu, peut-on généraliser la procédure à d'autres états mentaux que les croyances ?

Pour répondre à ces questions, plusieurs auteurs (Burge 1996 ; Bilgrami 2006 ; Moran 2001 ; Frankish 2004) proposent de partir de la conception que nous nous faisons de nous-mêmes comme penseurs et agents rationnels — une conception qui est au fondement de l'activité de raisonnement critique. Pour cette raison, nous appellerons ce point de vue la « position rationaliste ». Avant d'aller plus loin, soulignons qu'il est aujourd'hui fréquent, dans la littérature psychologique, de distinguer entre deux systèmes de raisonnement : le système 1, qui opère de façon automatique et rapide, sans demander d'effort d'attention, et sans que le sujet ait le sentiment d'un contrôle volontaire ; et le système 2, plus lent, qui repose sur l'attention, dépend de la volonté consciente, et dont les opérations sont au moins en partie accessibles au sujet (Stanovich 1999 ; Kahneman 2011). Dans toute la discussion qui va suivre, il sera question du système 2, c'est-à-dire du raisonnement accompagné de réflexion critique, d'investissement attentionnel, et d'expériences conscientes. Selon les défenseurs de la conception rationaliste de la connaissance de soi, raisonner de façon critique est une activité gouvernée par des normes, dont l'exercice implique donc des engagements, et cette activité est constitutive de ce que c'est, pour le comportement d'un agent, d'être causé par des attitudes propositionnelles assujetties aux normes de la rationalité.

En ce sens, articuler une croyance ou une décision, que ce soit d'ailleurs publiquement ou dans son for intérieur, ne revient ni à exprimer cette croyance, ni à décrire un état de son système cognitif, mais bien plutôt à formuler un engagement, théorique dans le cas de la croyance, ou pratique dans celui de la décision, vis-à-vis de l'attitude en question. Une source importante de l'approche rationaliste se trouve dans l'ouvrage classique d'Elizabeth Anscombe, L'intention (Anscombe 2000/1957). Il existe selon Anscombe un lien essentiel entre une forme de connaissance, qu'elle nomme « connaissance pratique », et un certain type d'application de la question « pourquoi ? » dirigée vers les actions : cette question n'attend de réponse portant sur des raisons d'agir que si l'agent possède une connaissance directe, immédiate, donc non fondée sur l'observation, de l'action qu'il accomplit. Lorsqu'il existe une description sous laquelle un comportement est intentionnel, il est possible d'identifier des raisons qui, du point de vue de l'agent, sont susceptibles de justifier le comportement relativement à une certaine fin recherchée. Considérer l'action comme gouvernée par une intention de l'agent, cela présuppose en effet que celui-ci soit capable de décrire celle-ci comme un moyen visant à la réalisation d'une certaine fin, et qu'il soit arrivé à

cette conclusion au travers d'une délibération pratique. Selon cette perspective, la connaissance qu'a l'agent de sa propre intention n'est pas une connaissance théorique fondée sur l'observation : il s'agit d'une connaissance pratique, fondée sur la connaissance des raisons qui rendent l'action raisonnable de son point de vue. Les rationalistes reprennent l'intuition centrale d'Anscombe : l'intention n'est pas un état passif de l'agent, mais la conclusion atteinte à l'issue d'une réflexion consciente. S'il se conçoit comme un être rationnel, l'agent doit donc considérer qu'il est engagé à réaliser le contenu de son intention. Supposons par exemple qu'Hoeradip arrive, à l'issue d'un tel raisonnement pratique, à la conclusion selon laquelle s'inscrire à la faculté de Médecine est la meilleure manière de réaliser le but consistant à devenir cardiologue. Connaître cette conclusion, cela revient dans le cadre rationaliste à connaître le contenu de son intention, qu'il exprimerait par l'aveu « j'ai l'intention de m'inscrire à la faculté de Médecine », et qui constitue un certain engagement dans l'action.

On voit donc que dans cette perspective, la connaissance réflexive qu'un agent forme de ses propres attitudes possède des caractéristiques qui la distingue des autres formes de connaissances. En premier lieu, c'est parce que nos attitudes sont constitutivement liées à nos raisons et nos engagements vis-à-vis des normes du raisonnement que nous pouvons les connaître avec une autorité spécifique. Le concept même d'agent rationnel implique que les attitudes de cet agent soient sensibles à des raisons, et puissent donc se modifier en fonction des données rationnelles disponibles. En nous inspirant d'un exemple de Tyler Burge (Burge 1996), supposons qu'on demande à Sherlock Holmes de justifier sa croyance que Moriarty a commis un certain crime. Pour répondre, le détective examinera non pas ses croyances, mais bien les données probantes qu'il a de croire que Moriarty est le coupable. Si l'examen des données ne lui semble plus justifier cette hypothèse, il devra, en tant qu'agent rationnel, la modifier, et donc modifier la croyance correspondante. La perspective rationaliste reprend donc l'intuition centrale de Gareth Evans — nous examinons nos raisons de croire afin de déterminer ce que nous croyons —, mais la complète par une idée importante : la possibilité même d'une connaissance privilégiée et autorisée de nos propres états repose sur le concept d'agent rationnel sensible à des normes. A cet égard on trouve chez Finkelstein l'une des formulations les plus claires de la conception rationaliste de la connaissance réflexive des croyances, qui montre très bien que l'on substitue, dans cette perspective, une question normative (« que dois-je croire ? ») à une question descriptive (« qu'est-ce que je crois » ?) :

« La question de savoir si je crois que P est, pour moi, le reflet de la question de savoir ce que je dois, rationnellement, croire — i.e. de la question de savoir s'il y a des raisons qui m'obligent à croire que P. Je peux répondre à la première question en répondant à la seconde » (Finkelstein 2012, 103).

On comprend donc mieux dès lors pourquoi la conception de l'introspection comme observation interne est inadéquate selon ces auteurs. Il est certes possible d'entretenir une relation d'observation détachée vis-vis de ses propres attitudes propositionnelles. (Moran 2001) comme (Burge 1996) prennent, à ce propos, l'exemple de la cure analytique, qui peut nous conduire à considérer de manière détachée certaines de nos croyances ou de nos motivations. Grâce à la cure, un patient peut découvrir certains de ses désirs cachés en adoptant une perspective interprétative sur ses propres comportements. Mais une telle perspective, que Moran qualifie d'« aliénée », est étrangère au concept d'attitude associé à la

pratique du raisonnement critique. Adopter la perspective d'un agent rationnel sur ses attitudes, c'est considérer qu'elles doivent être modifiées en fonction des données disponibles, et donc qu'elles sont sujettes à des normes qui impliquent une forme de responsabilité. Or, un tel agent ne peut se considérer comme responsable de croyances ou de désirs acquis indépendamment de la considération de raisons, et c'est en ce sens qu'il entretient une relation d'aliénation avec de tels états mentaux. La transparence au monde n'est donc pas une caractéristique de tous les états mentaux, mais uniquement de ceux pour lesquels une attitude critique, réflexive, est possible, ce que Moran explique clairement dans le texte suivant :

« (…) lorsque je me conçois moi-même comme un agent rationnel, la prise de conscience (awareness) de ma croyance est la prise de conscience du fait que je sois engagé vis-à-vis de sa vérité, un engagement qui transcende quelque description de mon état psychologique que ce soit. Et l'expression de cet engagement se manifeste par le fait que mes rapports de cette croyance sont soumis à la condition de transparence : considérer X (et rien d'autre que X) est une condition nécessaire à laquelle mon rapport de la croyance que X est soumis » (Moran 2001, 84).

Très influencé par Elizabeth Anscombe, Moran rejoint sur un point au moins l'approche expressiviste de la connaissance de soi : la transparence des attitudes est liée, selon lui, à notre capacité à les endosser, à nous engager rationnellement vis-à-vis de leurs contenus, qui se manifeste précisément dans les aveux. Dire « j'ai l'intention de m'inscrire en faculté de Médecine », ce n'est pas tant décrire un fait mental qu'exprimer un engagement personnel vis-à-vis du contenu de l'intention.

L'approche rationaliste explique-t-elle véritablement la nature de la connaissance que les sujets forment sur leurs propres états mentaux ? Soulignons d'abord que Burge, Moran et Bilgrami ne prétendent pas proposer une théorie complètement générale et unifiée de la connaissance de soi. La théorie ne s'applique qu'aux états liés à des engagements rationnels associés à des normes, donc aux attitudes propositionnelles sensibles aux raisons. Même si l'on se restreint à ces attitudes, la théorie proposée n'est cependant pas complètement claire, et il n'est pas évident non plus que tous les tenants du rationalisme aient une vision commune sur ce point (Gertler 2011). Un élément commun réside sans doute dans la conviction selon laquelle la connaissance de soi n'a pas besoin de se fonder sur des justifications renvoyant à des données probantes. Burge propose ainsi de compléter le concept de justification par celui d'autorisation (entitlement) : puisque le concept même de réflexion critique présuppose qu'un agent soit capable d'engagements vis-à-vis de ses attitudes, un tel agent pourra être dit autorisé à s'auto-attribuer une certaine attitude à partir du moment où il sera en position d'engager une telle réflexion critique sur son contenu. De fait, si l'on interprète les aveux comme exprimant des engagements, réclamer qu'ils soient justifiés par des données probantes reviendrait sans doute à commettre une erreur de catégorie : un engagement n'est pas le genre d'énoncé qui peut être justifié de la sorte (Burge 1996).

Même en adoptant ce concept d'autorisation, on peut se demander cependant si la théorie offre réellement une perspective entièrement nouvelle sur la connaissance de soi. En effet, un agent n'est autorisé, au sens de Burge, à s'auto-attribuer une attitude, qu'à partir du moment où il endosse le contenu de l'attitude dans un acte conscient, par exemple dans un

jugement conscient ou dans une décision consciente. Après tout, ce qui autorise l'agent à procéder à une auto-attribution de croyance, c'est qu'il lui apparaît consciemment qu'il est rationnellement engagé vis-à-vis du jugement correspondant. Autrement dit, la théorie présuppose que l'agent a un accès conscient direct au contenu du jugement — ou à une décision dans le cas des auto-attributions d'intentions. C'est un point problématique, comme le souligne Peter Carruthers (Carruthers 2011, 102). S'il est incontestable que nous ayons un accès conscient aux contenus d'énoncés formulés dans le discours intérieur, il ne va pas de soi qu'il existe un lien direct entre ces énoncés d'une part, et d'autre part nos décisions et nos jugements. Supposons qu'un agent s'entende dire, dans son discours intérieur : « c'est décidé, je m'inscris en médecine ». Selon la perspective rationaliste, il fait l'expérience directement et consciemment de l'acte mental consistant à se décider. Mais une autre interprétation, empiriquement plus plausible, est possible : qu'il fait l'expérience d'un acte cognitif d'énonciation intérieur, dont le lien causal avec un hypothétique acte de décision est sans doute très complexe — à supposer qu'il existe. On le sait bien : il n'est ni nécessaire ni suffisant de se dire intérieurement qu'on a pris une décision pour l'avoir réellement prise. Ces remarques remettent en question l'approche rationaliste. En effet, si l'on suppose que l'on n'a pas un accès direct, mais bien un accès interprétatif, à ses jugements et à ses intentions de premier ordre, l'idée centrale selon laquelle il existerait une dissymétrie fondamentale entre des états accessibles de manière détachée et aliénée d'une part, et d'autres états vis-à-vis desquels l'agent s'engagerait consciemment, se trouve fragilisée. Quassim Cassam a souligné une autre faiblesse de l'approche rationaliste (Cassam 2014). Cette approche repose sur la substitution de questions normatives (« que dois-je croire ? », « que dois-je décider ? ») aux questions descriptives (« qu'est-ce que je crois ? », « qu'est-ce que je préfère ? »). En cela, elle tente cependant d'éclairer un problème difficile (comment répondre aux questions descriptives par l'introspection ?) par un problème encore plus difficile (comment répondre aux questions normatives, et ceci sans faire usage du concept traditionnel d'introspection ?). Un problème qui, de plus, n'est essentiellement lié au premier que si l'on accepte le caractère adéquat de l'image de l'agent rationnel sur laquelle repose l'approche rationaliste. Si l'on doute que cette image soit autre chose qu'une idéalisation, la théorie rationaliste vaudra peut-être pour cette idéalisation, qu'il nomme ironiquement l'homo philosophicus, mais pas pour les homo sapiens que nous sommes (Cassam 2014, 27).

1. Connaissance des états sensoriels et rationalisme

Peut-on généraliser certaines des intuitions de la théorie rationaliste de la connaissance réflexive des attitudes à la connaissance d'autres états mentaux, en particulier celle des états sensoriels ? C'est ce que Christopher Peacocke propose dans une série de publications influentes (Peacocke 1999, 2008, 2014). Selon lui, une expérience consciente de perception constitue une raison suffisante d'une auto-attribution introspective. Il soutient que les agents rationnels suivent une règle qu'il nomme la « règle fondamentale », qui permet de passer d'un jugement de perception à l'auto-attribution correspondante. Le texte suivant illustre le fonctionnement de la « règle fondamentale » :

« Aristote soutenait que c'est par la vue que l'on perçoit que l'on est en train de voir. L'idée qui se trouve au coeur de la position d'Aristote me semble juste, à condition qu'on la comprenne de la façon suivante : c'est par la vue que l'on sait que l'on voit. Supposons que vous soyez en train de voir que :

(1) Ce bureau est couvert de documents.

Cette connaissance visuelle portant sur le monde vous donne une bonne raison de juger que l'auto-attribution suivante est correcte :

(2) Je vois que le bureau est couvert de documents.

Il s'agit d'une transition que vous êtes autorisé à faire, à partir d'un état conscient correspondant à un vécu vers un jugement. Si un penseur en vient à juger, de cette manière, qu'il voit que ce bureau est couvert de documents, son jugement peut donc être considéré comme une connaissance ». (Peacocke, 2008, 206-207)

Peacocke emprunte à Tyler Burge le concept d'autorisation, qui joue un rôle clef dans sa théorie. En effet, une expérience visuelle véridique portant sur un bureau couvert de documents ne peut certainement pas constituer la justification, au sens habituel, d'une auto-attribution. L'expérience porte en effet sur le monde extérieur : sur le bureau, les objets qui s'y trouvent, et sur leurs propriétés visibles. Or, pour qu'une expérience sensorielle dotée d'un certain contenu puisse justifier un état cognitif, il faut au minimum qu'il existe une relation d'implication logique entre le contenu de l'état sensoriel et celui de l'attitude. Ainsi, l'état visuel décrit par (1) peut être dit justifier la croyance selon laquelle il y a des documents sur le bureau, car le contenu intentionnel de l'état implique celui de la croyance. Si l'état est véridique, il est impossible que la croyance associée ne le soit pas. Néanmoins, Peacocke maintient qu'il existe bien une transition rationnelle entre l'expérience véridique et l'auto-attribution, quoique la rationalité de cette transition ne provienne pas d'une relation de justification, mais plutôt d'une relation d'autorisation. Pour comprendre cette position, il faut commencer par souligner que toutes les transitions rationnelles ne reposent pas sur la relation de justification (Peacocke, 2004). Supposons par exemple que de la prémisse « Amira a arrêté de fumer » on infère : « Amira fumait ». La proposition selon laquelle Amira a arrêté de fumer ne peut pas justifier la conclusion selon laquelle elle fumait, puisque la première proposition n'implique pas logiquement la seconde. Il existe cependant une relation de présupposition entre les deux jugements : juger qu'Amira a arrêté de fumer présuppose qu'elle fumait, et pour cette raison, on peut considérer que la conclusion est rationnelle, et donc qu'on est autorisé à l'affirmer. De manière comparable selon Peacocke, formuler un jugement perceptif du type « ce bureau est couvert de documents » présuppose que l'agent ayant jugé l'a fait sur la base d'une certaine expérience consciente. Par conséquent, celui-ci est immédiatement autorisé à s'auto-attribuer la croyance « je vois que le bureau est couvert de documents ».

La position de Peacocke est séduisante, car elle met la thèse de la transparence au coeur de la théorie des auto-attributions d'états perceptifs : seuls les états de premier ordre eux-mêmes constituent des raisons des auto-attributions. Néanmoins il n'est pas certain qu'elle puisse rendre compte de manière satisfaisante de la sécurité tout à fait spécifique des auto-attributions. Considérons en effet une situation dans laquelle un agent se trouve non pas dans un état perceptif, mais plutôt dans un état sensoriel non-véridique — qu'il s'agisse d'une hallucination, d'une illusion, ou d'un état d'imagination visuelle ou auditive. Il paraît clair qu'il suffit à l'agent de vivre consciemment l'expérience en question pour pouvoir se l'auto-attribuer. On peut former non seulement des connaissances réflexives portant sur nos états perceptifs, mais tout aussi bien des connaissances portant sur les états sensoriels non-

véridiques, et il semblerait étrange que la méthode qui nous permet de former les connaissances réflexives dans le premier cas ne puisse pas s'appliquer dans le second. Or l'application de la « règle fondamentale » proposée par Peacocke aux sensations non-véridiques semble difficile (Ludwig 2005). Comment, en effet, être dans un état non-véridique pourrait-il autoriser rationnellement quelque transition inférentielle que ce soit ? C'est un problème important, car il est essentiel à la connaissance introspective des sensations de pouvoir être acquise même dans des contextes d'hallucination ou d'illusion. Comme nous y insistions plus haut, le fait qu'une expérience soit illusoire ne nous empêche nullement de former une connaissance introspective des sensations qui la constituent.

6. Les approches inférentialistes

1. L'idée centrale de l'inférentialisme

Quoique sceptiques vis-à-vis de l'idée d'un accès privilégié à un monde mental intérieur, les rationalistes restent attachés à la thèse cartésienne selon laquelle la connaissance introspective est immédiate, donc non-inférentielle. En ce sens, ils appartiennent à la tradition ouverte par Descartes, dont Christopher Peacocke s'est d'ailleurs fait le défenseur (Peacocke 2012 ; 2014). Les philosophes qui, depuis (Ryle 2013/1949), se rattachent à ce que nous appellerons « inférentialisme », remettent en revanche en question le caractère immédiat de l'introspection. Il s'agit cette fois d'une rupture radicale avec Descartes, que cette citation de Gilbert Ryle résume très bien :

« Les sortes de choses que je peux découvrir sur moi-même sont les mêmes que les sortes de choses que je peux découvrir sur les autres personnes, et les méthodes qui permettent de les découvrir sont dans une large mesure les mêmes. (…) En principe sinon en pratique, les méthodes par lesquelles John Doe découvre des choses sur John Doe sont les mêmes que celles par lesquelles il découvre des choses sur Richard Roe» (Ryle 2013/1949, 165-166).

Ce que soutient Ryle dans ce texte, c'est que nous avons besoin d'interpréter notre propre comportement pour acquérir des informations sur nos propres états mentaux, exactement comme nous devons interpréter le comportement d'autrui. C'est ce qu'il faut entendre par « inférentialisme » : la connaissance de soi repose sur des inférences interprétatives fondées sur les données comportementales. Ces inférences causent les croyances introspectives, et elles les justifient, ce qui veut dire que le degré de jusfication d'une croyance introspective peut dépendre des données comportementales qui en constituent la base empirique (Cassam 2014).

L'inférentialisme, s'il a été défendu en particulier en psychologie depuis assez longtemps (Bem 1972 ; Gopnik 1993), se heurte de prime abord à de puissantes objections (Cassam 2014, 149). La première est que l'inférentialisme, puisqu'il met sur le même plan la connaissance introspective et la connaissance des autres esprits, semble renoncer à l'idée cartésienne d'une dissymétrie fondamentale entre ces deux types de connaissances. On peut même se demander s'il y a encore un sens, chez un auteur comme Ryle, à parler d'introspection à propos de la connaissance de soi. Une seconde objection très forte, c'est qu'il semble y avoir d'évidents contre-exemples à l'inférentialisme. N'est-il pas manifeste, ainsi, que je n'ai pas besoin d'observer mon propre comportement pour savoir que j'ai mal ? N'est-il pas manifeste que je peux sentir la colère monter en moi sans avoir pour cela à observer la déformation des traits sur mon visage qui l'accompagne ?

Quoique pressantes, ces objections ne sont pas décisives. En ce qui concerne la première objection, il faut souligner que l'inférentialisme peut parfaitement s'accomoder d'une différence de degré, sinon de nature, entre la connaissance introspective et la connaissance des autres esprits. C'est un point que Ryle lui-même soulignait déjà, et qui est repris par les inférentialistes contemporains (Lawlor 2009 ; Carruthers 2011 ; Byrne 2012 ; Cassam 2014). Considérons de nouveau l'exemple de la prise de conscience d'une émotion, par exemple de ma colère en train de monter. Je peux certainement prendre conscience de ma colère bien avant un observateur extérieur si je suis attentif, mais un inférentialiste peut l'expliquer : j'ai en effet accès à des informations, à des données, dont l'observateur extérieur ne dispose pas. Ainsi, le comportement de colère s'accompagne typiquement d'une augmentation du rythme cardiaque, qui n'est pas facilement détectable pour un observateur extérieur mais que le sujet peut percevoir. Notons que l'information portant sur l'augmentation de mon rythme cardiaque n'est pas une information introspective. Une philosophe inférentialiste peut donc maintenir que le sujet a accès de façon privilégiée à certaines informations, portant sur ses états corporels, sur les actions qu'il est en train d'accomplir, voir peut-être même sur ses actions mentales : comme nous le verrons plus bas, Peter Carruthers, reprenant une intuition de Ryle, considère ainsi que nous avons directement accès au contenu des énoncés lorsque nous nous parlons à nous-mêmes dans le discours intérieur, et que nous pouvons inférer certains de nos états mentaux à partir de cette connaissance directe (Carruthers 2011 ; Byrne 2012). Par ailleurs, on peut inverser l'argument qui sous-tend l'objection : n'est-il pas possible que nous soyons, dans certaines circonstances, moins bien placées que les observateurs extérieurs pour identifier nos états mentaux ? C'est ce que soutient Ryle, et nous verrons que les données portant sur la confabulation lui donnent largement raison.

N'est-il cependant pas incontestable que nous avons un accès non-inférentiel

aux contenus des états perceptifs ou sensoriels, par exemple aux contenus des expériences de douleur ? Plusieurs inférentialistes suggèrent qu'il s'agit en effet d'authentiques contre-exemples, c'est-à-dire de connaissances qui ne reposent pas sur des inférences interprétatives. Ils soulignent néanmoins que l'on peut être inférentialiste vis-à-vis d'un type d'introspection, sans l'être vis-à-vis de tous les types d'introspection. Du point de vue phénoménologique, il semble y avoir peu de rapport entre la connaissance introspective de la croyance selon laquelle il y a des montagnes sur la Lune, la connaissance introspective de la décision de sortir acheter du lait, ou la connaissance introspective d'une sensation de douleur. Selon Carruthers, seule la dernière connaissance repose sur un mécanisme non-inférentiel (Carruthers 2011). A cet égard, les philosophes inférentialistes insistent sur la diversité et la richesse de la connaissance introspective (Cassam 2014). Dans les développements qui suivent, nous allons présenter le traitement inférentialiste de la connaissance introspective de différentes sortes d'états mentaux : les états dispositionnels tels que les préférences ou les croyances, les états occurrents mais non sensoriels comme les jugements ou les décisions, puis les états sensoriels.

1. **Introspection et confabulation**

Nous nous concentrerons dans cette partie sur la connaissance introspective des états mentaux dispositionnels : ces états mentaux qui ne sont pas occurrents au moment auquel le sujet exerce sa capacité d'introspection pour les connaître. Un état mental dispositionnel n'est pas un évènement avec un début et une fin. Nous opposerons donc les états mentaux

dispositionnels aux états mentaux occurrents, ou évènements mentaux. Au contraire d'un état mental dispositionnel, un évènement mental peut avoir lieu à un certain moment dans le temps : me souvenir de mes vacances, imaginer ma maison, faire du calcul mental, décider d'aller prendre le bus ou juger que les poulpes sont des animaux passionnants sont autant d'évènements mentaux. Savoir par introspection si je suis bien en train de juger que la peine de mort est injuste, c'est exercer l'introspection sur un évènement mental. Savoir si je suis en faveur ou en défaveur de la peine de mort par introspection, c'est exercer l'introspection sur un état mental dispositionnel : la croyance n'est pas, pour ainsi dire, active au moment de l'introspection.

Depuis une quarantaine d'années, la conception inférentialiste de l'introspection des états mentaux dispositionnels est confirmée par les données de la psychologie sociale. La conception inférentialiste de l'introspection considère que nous n'avons qu'un accès indirect (inférentiel), ou interprétatif, à nos états mentaux dispositionnels, comme nos croyances, nos désirs, nos buts ou nos intentions. L'introspection de ces états, selon les inférentialistes, serait davantage une capacité d'inférence à partir d'un ensemble d'indices comportementaux qu'une voie d'accès directe et fiable à nos états mentaux. Dans un article célèbre, Nisbett et Wilson avancent que l'introspection nous amène généralement à en dire « plus que ce que nous savons », ou pouvons savoir, sur nos états mentaux dispositionnels (Nisbett et Wilson 1977). Dans leur protocole expérimental, les deux expérimentateurs disposaient de gauche à droite quatre paires de collants (de gauche à droite : A, B, C, D) exactement identiques et demandaient à des consommateurs de les évaluer. L'hypothèse de Nisbett et Wilson était que les participants préfèrent systématiquement les collants qui se trouvent le plus à droite (la dernière à être au centre de l'attention des participants), et cette hypothèse s'est trouvée validée. Les sujets choisissaient la paire de collants A (la plus à gauche) à 12%, B à 17%, C à 31% et D à 40 %. Par ailleurs, ils ne remarquaient pas, dans cette expérience, que toutes les paires étaient identiques et que la large préférence pour la paire D était simplement une préférence due à la position des collants. À la suite de leur choix, il était demandé aux participants d'expliquer les raisons qui les conduisaient à préférer la paire D à une autre paire de collants. Ces derniers justifiaient leur choix en expliquant aux expérimentateurs que la paire de collants D était plus élastique, ou plus douce, mais aucun ne mentionnait que la position avait joué un rôle quelconque dans leur préférence. Lorsqu'on leur demandait s'ils pensaient que la position des collants avait influencé leur décision, tous les participants répondaient que ce n'était pas le cas. Les participants justifiaient donc leur choix par des qualités qui n'existaient pas effectivement et qui, par conséquent, n'avaient pas pu les conduire à un tel choix. L'introspection sur les raisons de leurs préférences menait donc les sujets à s'attribuer un jugement qui n'avait aucun rôle causal effectif dans la formation de ces préférences. Au contraire, la différence qui avait effectivement jouée un rôle dans le choix de la paire de collants, c'est-à-dire leur position, n'était jamais mentionnée par les participants. Nous constatons donc, grâce à cette expérience, que la manière dont nous expliquons nos actions peut ne pas correspondre aux causes effectives de l'action. Par conséquent, l'introspection peut nous induire en erreur, à en dire plus que ce que nous savons, ou à confabuler.

L'introspection mène les sujets à justifier leurs choix par des propriétés inexistantes. La confabulation peut être définie comme un mensonge honnête, dont le premier convaincu est le menteur lui-même. Ce phénomène est un symptôme dans plusieurs pathologies et

apparaît, à l'origine, chez les patients atteints du syndrome de Korsakoff. Les patients atteints de ce syndrome sont amnésiques et confabulent des souvenirs : lorsqu'il est demandé à un patient ce qu'il a fait le jour précédent, ce dernier répond qu'il était « au bureau toute la journée » alors même qu'il est à la retraite depuis plus de quinze ans (Hirstein 2000). La confabulation apparaît aussi dans les cas d'anosognosie, dans lesquels les patients nient une pathologie. Un patient paralysé pourra ainsi répondre qu'il refuse de bouger du fait de son arthrite, et ce en toute bonne foi. L'anosognosie conduit les sujets à confabuler, de telle sorte qu'ils n'auront jamais conscience de l'existence de leur maladie, aussi grave soit-elle.

Si la confabulation a d'abord été observée dans des cas pathologiques, elle est devenue de plus en plus étudiée chez les sujets sains. Ainsi, dans une étude menée par Johansson et al. (Johansson et al. 2005), les expérimentateurs montraient des photographies de visages à des sujets qui devaient choisir, parmi deux propositions, le visage qu'ils préféraient et expliquer les raisons pour lesquelles ils avaient choisi cette photo. Mais les expérimentateurs, subrepticement, échangeaient la photo choisie par les sujets avec la photo qui n'avait pas été retenue : non seulement les sujets ne s'apercevaient pas, dans leur immense majorité, que les photos avaient été échangées (alors même qu'elles étaient parfois ouvertement différentes), mais ceux-ci justifiaient leur « choix » en confabulant des raisons. Les sujets avançaient, par exemple, « j'ai choisi cette photo car j'aime bien les boucles d'oreilles de cette femme », alors que, sur la photo qu'ils avaient effectivement choisie, l'individu ne portait pas de boucles d'oreilles. Il suffisait donc aux expérimentateurs de faire croire aux sujets qu'ils avaient choisi une photographie pour que les sujets justifient ce qu'ils pensaient être leur choix en confabulant. De même que dans l'étude de Nisbett et Wilson, la justification apportée par les sujets est indépendante de ce qui a effectivement motivé le choix en premier lieu. L'introspection sur les causes de nos choix est donc faillible, et bien plus qu'une voie d'accès à nos motifs d'actions, elle se révèle avoir un usage essentiellement justificatif. Cette thèse inférentialiste quant à l'introspection est défendue principalement par des psychologues sociaux comme Timothy Wilson (Wilson 2002) : l'introspection n'offre pas un accès direct aux causes de nos comportements, bien plutôt, elle n'est qu'une façon d'interpréter notre comportement pour en déterminer les causes post facto. Afin de nous attribuer des états mentaux dispositionnels comme des croyances, des désirs, des buts et des intentions, notre utilisation de l'introspection reposerait bien plus sur une interprétation de notre comportement que sur un accès direct à ces états mentaux dispositionnels.

La position inférentialiste est aussi confirmée par les phénomènes « d'auto-perception », principalement théorisés par le psychologue social Daryl Bem (Bem 1972) et dont un exemple paradigmatique a été réalisé dans une expérience de Wells et Petty (Wells et Petty 1980). Leur expérience consistait à demander aux sujets de hocher la tête de haut en bas tandis qu'ils écoutaient un discours à travers des écouteurs. Les participants, qui pensaient alors que l'étude n'était qu'un test de la fiabilité des écouteurs, étaient ensuite interrogés sur leur degré d'approbation au discours qu'ils avaient écouté durant l'essai. Les résultats comparés à un groupe inactif ou hochant la tête de gauche à droite montraient que les sujets auxquels il était demandé de hocher la tête de haut en bas durant l'essai étaient nettement plus convaincus par le discours que le groupe de contrôle. Une fois de plus, l'introspection a un résultat contre-intuitif. Si l'introspection était pour les sujets une voie d'accès directe à leurs croyances, pourquoi le mouvement de tête aurait-il influencé le degré de conviction des participants ? L'hypothèse du camp inférentialiste est que, lors de l'introspection, les sujets se perçoivent

eux-mêmes hocher la tête et interprètent ce comportement pour en conclure qu'ils doivent être d'accord avec le message qui leur est diffusé. Contrairement à ce qui pourrait être prédit par une théorie postulant un accès non-inférentiel à nos attitudes, dans ces expériences, aucun sujet n'a jamais rapporté l'influence du hochement de tête sur le degré de croyance. L'étude a été répliquée à de nombreuses reprises, sous différents paradigmes. Brinol et Petty (Brinol et Petty 2003) ont demandé à des sujets d'écrire trois choses sur eux-mêmes, positives ou négatives, susceptibles d'avoir un impact sur leurs vies professionnelles. Les sujets devaient accomplir cette tâche soit de leur main forte, soit de leur main faible. Les sujets dans la condition « main faible » exprimaient systématiquement moins de confiance en eux. Loin d'offrir une connaissance directe de nos propres croyances, l'introspection se baserait donc au moins en partie sur une interprétation de nos propres comportements lorsqu'il est question de nous attribuer à nous-mêmes des états mentaux dispositionnels tels que des croyances. Comme le souligne Quassim Cassam, les travaux de Bem et leurs prolongements sont philosophiquement importants, car ils montrent que même si l'on admet que les données commportementales n'ont pas forcément de pertinence pour les auto-attributions, elles peuvent incontestablement être pertinentes dans certaines situations (Cassam 2014, 148).

Nous avons vu que la psychologie sociale offre des données suggérant une approche inférentialiste de l'introspection des états mentaux dispositionnels. Dans la prochaine section, nous explorons la possibilité d'une approche inférentialiste de l'introspection des évènements mentaux comme les décisions ou les jugements (Carruthers 2009, 2011, 2013). Dans la section suivante, nous montrerons qu'il est même possible d'entretenir une position inférentialiste, ou sceptique, quant à l'introspection de nos états mentaux perceptifs occurrents (Schwitzgebel 2008, 2011, 2012).

1. Introspection et attitudes propositionnelles occurrentes

L'introspection est censée garantir un accès privilégié à nos propres attitudes propositionnelles. Une attitude propositionnelle, comme le nom l'indique, est une attitude possible envers une proposition. Il est possible d'avoir de nombreuses attitudes différentes envers la proposition « il pleut ». Ainsi, je peux croire qu'il pleut, penser qu'il pleut, souhaiter qu'il pleuve, détester lorsqu'il pleut, dire qu'il pleut, et ainsi de suite. Toutes ces attitudes sont appelées « attitudes propositionnelles ». Avoir des intentions, des croyances, des buts, des désirs, c'est avoir des attitudes propositionnelles. Les attitudes propositionnelles peuvent, à un certain moment, être occurrentes : c'est le cas lorsque je juge qu'il pleut, lorsque je décide de me rendre à la banque, ou lorsque je désire activement une seconde part de gâteau.

D'après les théories non-inférentialistes de l'introspection, alors que je dois interpréter le comportement d'autrui pour en déduire ses attitudes propositionnelles comme par exemple ses croyances, ses buts ou ses intentions, je dispose d'un accès direct à mes propres attitudes occurrentes par la voie de l'introspection. En d'autres termes, grâce à l'introspection il n'y aurait nul besoin d'interpréter les données portant sur mon propre comportement pour savoir ce que je juge ou ce que je veux. Selon cette théorie traditionnelle de l'introspection, l'accès dont je dispose à mes propres attitudes propositionnelles est d'un type fondamentalement différent de mon accès aux attitudes d'un autre individu. L'introspection instaurerait donc une dissymétrie entre la connaissance de soi et la connaissance d'autrui : d'un coté un accès privilégié, de l'autre, un accès interprétatif.

D'après les données présentées dans la section précédente, il est possible de douter de la validité de cette théorie traditionnelle de l'introspection lorsqu'il est question de nos attitudes propositionnelles non-occurrentes (ou dispositionnelles). Il semble en effet difficile d'expliquer la confabulation dans un tel cadre théorique : pourquoi les sujets se tromperaient-ils sur leurs propres attitudes propositionnelles en confabulant s'ils disposaient réellement d'un accès direct à leurs attitudes propositionnelles ? La réponse commune des théories traditionnelles (Peacocke 2008 ; Gazzaniga 2000 ; Siewert 1998) est d'avancer que ces données prouvent simplement que nous n'avons pas un accès introspectif aux causes de nos comportements. La psychologie sociale, si elle prouve effectivement que nous interprétons notre comportement pour nous auto-attribuer des états mentaux dispositionnels, ne prouve pas que nous ne disposons jamais d'un accès direct à nos attitudes propositionnelles au moment où celles-ci se manifestent dans des évènements mentaux (dans des décisions, ou des jugements par exemple). Si une conclusion aussi forte ne peut pas être dérivée des expériences de psychologie sociale, il demeure qu'elles prouvent que l'interprétation joue un rôle insoupçonné dans l'auto-attribution d'attitudes propositionnelles. La position inférentialiste peut-elle être justifiée même dans le cas où nous prenons connaissance d'évènements mentaux par introspection ? Autrement dit, se pourrait-il que nous ne connaissions ce que nous sommes en train de juger, de désirer, ou de décider que par l'intermédiaire d'un processus introspectif reposant sur des inférences ?

Comme nous l'avons souligné plus haut, certains auteurs inférentialistes pensent que notre accès à certains évènements mentaux est inférentiel — par exemple pour Wegner, c'est le cas pour notre connaissance de nos intentions occurrentes (Wegner 2002). Peter Carruthers a également développé une théorie inférentialiste de l'introspection des attitudes propositionnelles représentant une synthèse des différentes approches de l'introspection en sciences cognitives et en psychologie sociale (Carruthers 2011). La théorie de Carruthers nie qu'il y ait une différence de nature entre l'accès aux attitudes propositionnelles en première et en troisième personne. En d'autres termes, selon Carruthers, l'introspection n'offre pas d'accès privilégié à nos propres attitudes propositionnelles, et ce, même lorsque celles-ci sont occurrentes. Il défend que le système cognitif responsable de nos capacités d'attribution d'états mentaux à autrui, le système de théorie de l'esprit (mindreading), est le même que celui qui nous permet de nous auto-attribuer des attitudes. Lorsque nous pratiquons l'introspection, nous ne faisons en réalité que tourner nos capacités de théorie de l'esprit sur nous-mêmes et nous auto-interpréter à la façon dont nous interprétons le comportement des autres. Si nous n'avons pas d'accès non-inférentiel à nos attitudes propositionnelles occurrentes, nous avons tout de même un accès direct à nos contenus sensoriels conscients : discours intérieur, imagination, émotions, perceptions. Lorsque nous appliquons notre théorie de l'esprit à nous-mêmes, le système de théorie de l'esprit n'a pas seulement accès à notre comportement, mais aussi à toutes ces sources d'informations sensorielles. C'est, selon Carruthers, la raison pour laquelle nous en savons davantage sur nos propres attitudes propositionnelles occurrentes que sur celles d'autrui malgré l'absence de différence de nature entre le type d'accès que nous avons à nos propres attitudes et à celles d'autrui. Cela explique aussi que nous pouvons continuer de nous auto-attribuer des attitudes en l'absence de tout comportement apparent. En effet, dans ces cas-là, notre conscience demeure un flux continu d'images et de discours intérieurs qui constituent une ressource d'informations permettant au système de théorie de l'esprit de garantir l'auto-attribution d'attitudes même en l'absence de comportement public.

Que se passe-t-il, selon Carruthers, lorsque, par exemple, je prends conscience par introspection de ma décision d'aller prendre le bus ? Premièrement, un évènement mental est créé : la décision d'aller prendre le bus. Deuxièmement, cette décision elle-même peut engager des systèmes perceptifs qui vont entraîner l'apparition d'un second évènement mental, par exemple, l'utilisation du discours intérieur pour me dire à moi-même consciemment « je vais aller prendre le bus ». Troisièmement, ce discours intérieur est interprété grâce à un système de théorie de l'esprit. Quatrièmement, à partir de cette interprétation, le système de théorie de l'esprit crée un jugement selon lequel j'ai décidé d'aller prendre le bus. Ce jugement peut à son tour engager des systèmes perceptifs entraînant l'apparition d'un nouvel évènement mental, par exemple, me dire à moi-même « j'ai décidé d'aller prendre le bus ». Le processus introspectif qui me mène d'un évènement mental, la décision, au jugement introspectif selon lequel j'ai décidé de prendre le bus, est interprétatif ou inférentiel. Je ne fais qu'inférer que je vais prendre le bus à partir des données accessibles, par exemple à partir d'un certain discours intérieur. Le système de théorie de l'esprit n'accède à aucun moment à la décision elle-même, tout ce à quoi il accède est un évènement mental sensoriel (le discours intérieur), qui est ensuite interprété comme étant le signe qu'une décision a été prise. Nous sommes ici bien loin d'une sorte de sens interne qui nous permettrait de conclure directement, à partir de la perception d'une décision par l'œil de l'esprit, qu'une décision a été prise.

Les données de psychologie sociale sur l'introspection et la confabulation, si elles ne permettent peut-être pas encore de justifier totalement une approche inférentialiste, ont au moins imposé une place pour l'interprétation dans le débat sur la nature de l'introspection des attitudes propositionnelles et évènements mentaux, de sorte que toute théorie de l'introspection doit désormais intégrer une composante interprétative. Cela rend la tâche plus difficile pour les théories non-inférentialistes de l'introspection, les vouant au mieux à intégrer deux voies d'accès à nos attitudes propositionnelles, l'une, directe, et l'autre, interprétative, à la façon de (Nichols et Stich 2003). Un nouveau problème se pose alors : les théories non-inférentialistes doivent justifier que des sujets disposant de capacités leur permettant un accès transparent à leurs attitudes propositionnelles procèdent parfois à une interprétation de leur propre comportement.

1. Introspection et expériences conscientes

Lorsque, pour certains inférentialistes comme Carruthers, nous n'avons pas d'accès direct à nos attitudes propositionnelles mais disposons d'une connaissance introspective directe de nos contenus sensoriels conscients, d'autres, comme Eric Schwitzgebel (Schwitzgebel 2011 ; Schwitgebel 2012), considèrent que, même dans ce cas, il y a lieu d'être sceptique. Ici, il n'est plus question de savoir si nous pouvons avoir un accès non-inférentiel à nos propres croyances, intentions, jugements, décisions et autres attitudes propositionnelles ou évènements mentaux. Plutôt, nous considérons la question de savoir si nous avons ou non un accès direct et infaillible, ou plus modestement, fiable, de nos propres états mentaux sensoriels conscients.

Wittgenstein écrivait : « Cause principale des maladies philosophiques ? Un régime unilatéral : on nourrit sa pensée d'une seule sorte d'exemples » (Wittgenstein 1953, § 593). La première critique de Schwitzgebel pourrait être parfaitement résumée par cet aphorisme appliqué à la littérature philosophique sur l'introspection. Les philosophes ont souvent alimenté la

discussion sur l'introspection d'exemples comme : « voir la couleur rouge », « ressentir une vive douleur » afin de prouver que l'introspection est susceptible de nous apporter une connaissance infaillible de notre expérience consciente. Il est couramment admis, même parmi les philosophes les plus sceptiques quant à nos capacités d'introspection (Dennett 2002), que nous ne pouvons nous tromper à propos de ce genre de phénomènes : lorsque nous jugeons que nous voyons du rouge dans de bonnes conditions, ou que nous ressentons une douleur, le jugement est inévitablement véridique. Mais comme l'avance Schwitzgebel, il semble nécessaire, avant de conclure à la fiabilité de l'introspection, de nous demander si ces exemples sont véritablement paradigmatiques. Le débat sur l'introspection ne serait-il pas atteint d'une « maladie philosophique » à ne se nourrir que « d'une seule sorte d'exemples » ? Nous considérons les cas comme « voir du rouge » ou « ressentir une douleur » et généralisons à partir de là comme sur des cas typiques. Schwitzgebel propose de partir de cas concrets et plus complexes. Nous pouvons déplacer le débat : « pourquoi ne pas prendre un autre point de départ pour changer ? » (Schwitzgebel 2003, 3).

Il y a d'innombrables questions à propos de nos expériences conscientes dont la réponse semble bien plus indéterminée que celles répondant à la question de savoir si nous ressentons de la douleur ou non. Où s'arrête notre champ de vision exactement ? Peut-on voir la couleur à sa périphérie ? Et si nous imaginons un paysage, les couleurs sont-elles vives, sont-elles même présentes ? Quel est le niveau de détail de cette image mentale ? De même, les émotions ont des contours très mal définis ; en éprouve-t-on toujours ? Peut-on comparer leurs intensités ? Ai-je toujours une humeur particulière ? Ai-je une idée claire de ce que c'est pour moi que d'être irritable, ou à l'aise ? Le débat sur la richesse de notre expérience consciente relève aussi de ces perplexités introspectives : ai-je toujours la sensation de mes pieds dans mes chaussures, ou cette sensation n'est-elle présente que lorsque j'y porte attention ? Si toutes ces questions peuvent probablement recevoir des réponses, nous devons toutefois reconnaître ici que l'introspection est un outil bien plus incertain pour y répondre que dans les exemples typiques invoqués en philosophie. Il est remarquable, écrit Schwitzgebel, que les propriétés que nous reconnaissons à nos expériences privées soient bien plus incertaines que les propriétés des objets extérieurs. Pourtant, c'est de la véracité de notre expérience du monde extérieur qu'a traditionnellement douté la philosophie et non de la fiabilité de notre expérience introspective. En réalité, Schwitzgebel montre que les qualités des objets externes sont bien plus aisément décidables que les qualités des objets privés.

Une controverse dans l'histoire de la psychologie offre un exemple d'indécision introspective. Après la création de la psychologie scientifique par Wundt, une série de psychologues utilisaient l'introspection des sujets et leurs rapports verbaux comme matière première de l'étude de phénomènes de haut niveau comme la mémoire ou l'imagination (Sackur, 2009). La psychologie introspectionniste se heurte néanmoins rapidement aux limites de sa méthodologie avec la « controverse de la pensée sans images » entre Titchener et Külpe (Lidenfeld 1978). Lorsque les sujets entraînés à l'introspection des expériences de Titchener rapportent pouvoir penser sans contenu sensoriel, les sujets menés par Külpe rapportent au contraire que toute pensée est basée sur un contenu sensoriel. L'introspection montre ici sa limite en tant que méthodologie pour une science psychologique, mais aussi comme moyen de connaître des faits basiques à propos de nos vies intérieures. L'existence même de cette querelle et son rôle dans l'histoire de la psychologie pourraient indiquer la pertinence de la position philosophique de Schwitzgebel sur l'introspection.

En considérant ce type d'exemples, l'introspection pourrait donc se révéler moins fiable qu'elle le semble lorsque nous considérons des cas simples. L'introspection pourrait même nous conduire à de fausses conclusions : contrairement à ce que nous dit notre expérience quotidienne, nous ne voyons pas la couleur aux extrémités de notre champ de vision, et notre expérience, visuelle notamment, est bien plus dépendante de notre attention qu'elle ne l'apparaît, comme en témoigne le phénomène de cécité inattentionnelle. L'introspection n'est donc pas une source de connaissance détaillée nous permettant de nous connaître infailliblement en première personne ; bien plutôt, l'analyse de Schwitzgebel suggère que notre accès à notre propre expérience consciente est moins fiable que nous le croyons d'ordinaire en nous limitant à des exemples simples.

En conséquence de ces différentes thèses inférentialistes et sceptiques, la psychologie semble pouvoir nous en apprendre autant sur notre expérience consciente que l'étude solitaire, introspective, de notre propre esprit, sinon d'avantage.

_____

BIBLIOGRAPHIE

Anscombe, G E M. 2000/1957. Intention. Cambridge, Mass.: Harvard University Press.

Armstrong, D M. 1968. A Materialist Theory of the Mind. London; New York: Routledge & K. Paul Humanities Press.

Aydede, Murat. 2005. Pain : New Essays on Its Nature and the Methodology of Its Study. Cambridge, Mass.: MIT Press.

Ayer, Alfred Jules. 1936. Language, Truth and Logic. London: Victor Gollancz Ldt.

Baars, Bernard J. 1993. A Cognitive Theory of Consciousness. Cambridge ; New York: Cambridge University Press.

Baddeley, Alan. 2003. Working memory: Looking back and looking forward. Nature Reviews Neuroscience 4 (10): 829-839.

Balog, Katalin. 2012. In defense of the phenomenal concept strategy. Philosophy and Phenomenological Research 84 (1): 1-23.

Bar-On, Dorit. 2004. Speaking My Mind : Expression and Self-knowledge. Oxford: Clarendon Press ; New York : Oxford University Press.

Bem, Daryl. 1972. Self-Perception theory. Advances in Experimental Social Psychology 6:1-62.

Bermúdez, José Luis (2010), Thought, Reference, and Experience : Themes From the Philosophy of Gareth Evans, Oxford, Oxford Clarendon Press.

Bilgrami, Akeel. 2006. Self-knowledge and Resentment. Cambridge, Mass.: Harvard University Press.

Block, Ned. 2009. Comparing the major theories of consciousness. In The Cognitive Neuroscience. Cambridge: Mass.: MIT Press.

Brinol, P et R Petty. 2003. Overt head movements and persuasion: A self-validation. Journal of Personality and Social Psychology 39:1123-1139.

Burge, Tyler. 1996. Our entitlement to self-knowledge: I. In Proceedings of the Aristotelian Society New Series, Vol. 96: 91-116.

Carruthers, Peter. 2007. Higher-order theories of consciousness. The Stanford Encyclopedia of Philosophy (Fall 2011 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2011/entries/consciousness-higher/>.

———. 2009. How we know our own minds: The relationship between mindreading and metacognition. The Behavioral and Brain Sciences 32 (2): 121-38; discussion 138–82.

———. 2011. The Opacity of Mind: An Integrative Theory of Self-knowledge. Oxford ; New York: Oxford University Press.

———. 2013. Mindreading the self. In Understanding Other Minds: Perspectives From Developmental Social Neuroscience. Ed. Simon Baron-Cohen, Michael Lombardo, and Helen Tager-Flusberg. Oxford University Press.

———. 2015. The Centered Mind : What the Science of Working Memory Shows Us About the Nature of Human Thought. New York: Oxford University Press.

Cassam, Quassim. 2014. Self-knowledge for Humans. Oxford: Oxford University Press.

Chalmers, David. 2003. The content and epistemology of phenomenal belief. in Q. Smith and A. Jokic (eds), Consciousness: New Philosophical Perspectives (Oxford, 2003), 220-271.

Churchland, Paul M. 1988. Matter and Consciousness : A Contemporary Introduction to the Philosophy of Mind. Cambridge, Mass.: MIT Press.

Conee, Earl. 1994. Phenomenal knowledge. Australasian Journal of Philosophy 72 (2): 136-150.

Crane, Tim. 2003. The intentional structure of consciousness. in Q. Smith and A. Jokic (eds), Consciousness: New Philosophical Perspectives (Oxford, 2003)

Dehaene, Stanislas and Lionel Naccache. 2001. Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. Cognition 79 (1): 1-37.

Dennett, Daniel. 2002. How could I be wrong? How wrong could I be? Journal of Consciousness Studies 9 (5-6).

Descartes, René, Michelle Beyssade, and Jean-Marie Beyssade. 2011. Méditations métaphysiques ; objections et réponses : suivies de quatre lettres. Paris: Flammarion.

Dretske, Fred. 2005. The epistemology of pain. in M. Aydede (dir.), Pain: New Essays on Its Nature and the Methodology of Its Study, Cambridge, Mass.: MIT Press, 59-73.

Dretske, Fred I. 1995. Naturalizing the Mind. Cambridge, Mass.: MIT Press.

Evans, Gareth. 1982. The Varieties of Reference. Oxford: Clarendon Press ; New York : Oxford University Press.

Falvey, Kevin. 2000. The basis of first-person authority. Philosophical Topics 28 (2): 69-99.

Finkelstein, David H. 2003. Expression and the Inner. Harvard University Press.

———. 2010. Expression and avowal. Wittgenstein: Key Concepts185-198.

———. 2012. From transparency to expressivism. In J. Conant et G. Abel (dir.), Rethinking Epistemology, vol. 2, Berlin et Boston, De Gruyter.

Frankish, Keith. 2004. Mind and Supermind. Cambridge: Cambridge University Press.

Fumerton, Richard A. 1995. Metaepistemology and Skepticism. Lanham, Md.: Rowman & Littlefield.

Gazzaniga, Michael S and Emilio Bizzi. 1995. The Cognitive Neurosciences. Cambridge, Mass.: MIT Press.

Gazzaniga, Mickael S. 2000. The Mind's Past. University of California Press.

Geach, Peter T. 1965. Assertion. The Philosophical Review, 449-465.

Gertler, Brie. 2001. Introspecting phenomenal states. Philosophy and Phenomenological Research 63 (2): 305-328.

———. 2011. Self-knowledge. New York: Routledge.

Goldman, Alvin I. 1979. What is justified belief? In G. Pappas G. (dir.), Justification and Knowledge. Boston: Reidel.

———. 2006. Simulating Minds : The Philosophy, Psychology, and Neuroscience of Mindreading. Oxford; New York: Oxford University Press.

Gopnik, Alison. 1993. How we know our minds: The illusion of first-person knowledge of intentionality. Behavioral and Brain Sciences 16 (01): 1.

Güzeldere, Güven. 1995. Is consciousness the perception of what passes in one's own mind? Conscious Experience335-357.

Harman, Gilbert. 1990. The intrinsic quality of experience. Philosophical Perspectives31-52.

Hatfield, Gary. 2002. Routledge Philosophy Guidebook to Descartes and the Meditations. London, Routledge.

Hill, Christopher S. 2009. Consciousness. Cambridge, UK; New York: Cambridge University Press.

Hirstein, William. 2000. Self-Deception and confabulation. Philosophy of Science 67:S418-S419.

Irvine, Elizabeth. 2013. Consciousness As a Scientific Concept : A Philosophy of Science Perspective. Dordrecht: Springer.

Johansson, Petter, Lars Hall, Sverker Sikstrom, and Andreas Olsson. 2005. Failure to detect mismatches between intention and outcome in a simple decision task. Science 310 (116): 116-9.

Kahneman, Daniel. 2011. Thinking, Fast and Slow. Farrar: Straus and Giroux.

Kant, Emmanuel. 1786. Premiers principes métaphysiques d'une science de la nature. Paris : Vrin. 1990.

Kind, Amy. 2003. Shoemaker, self–blindness and moore's paradox. The Philosophical Quarterly 53 (210): 39-48.

Kripke, Saul A. 1980. Naming and Necessity. Oxford: Blackwell.

Lawlor, Krista. 2009. Knowing what one wants. Philosophy and Phenomenological Research 79 (1): 47-75.

Leibniz, Gottfried Wilhelm. 1966. Nouveaux Essais Sur L'entendement Humain. Paris: Garnier-Flammarion.

Li, Wen, Isabel Moallem, Ken A Paller, and Jay A. Gottfried. 2007. Subliminal smells can guide social preferences. Psychological Science 18 (12): 1044-1049.

Lindenfeld, D.1978. Oswald Külpe and the Würzburg School. Journal of History of the Behavioral Sciences, 14, 132–141.

Locke, John and Pauline Phemister. 2008. An Essay Concerning Human Understanding. Oxford ; New York: Oxford University Press.

Lormand, Eric. 1996. Inner sense until proven guilty. Manuscrit.

Ludwig, Pascal. 2005. Une défense hétérodoxe de la conception inférentialiste de l'introspection. Dialogue 44 (01): 123-144.

Lycan, William G. 1996. Consciousness and Experience. Cambridge, Mass.: MIT Press.

McGinn, Colin. 1982. The Character of Mind. Oxford; New York: Oxford University Press.

Moran, Richard. 2001. Authority and Estrangement : An Essay on Self-knowledge. Princeton, N.J.: Princeton University Press.

Naccache, Lionel. 2006. Le Nouvel Inconscient : Freud, Christophe Colomb Des Neurosciences. Paris: O. Jacob.

Naselaris, Thomas, Kendrick N Kay, Shinji Nishimoto, and Jack L Gallant. 2011. Encoding and decoding in fmri. Neuroimage 56 (2): 400-410.

Nichols, Shaun and Stephen P Stich. 2003. Mindreading: An Integrated Account of Pretence, Self-awareness, and Understanding Other Minds. Oxford; Oxford; New York: Clarendon Press/Oxford University Press.

Nisbett, Richard and Stanley Schachter. 1966. Cognitive manipulation of pain. Journal of Experimental Social Psychology 2:227-236.

Nisbett, Richard E and Timothy D Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. Psychological Review 84 (3): 231-259. http://repositorium.uni-muenster.de/document/miami/7de7cfb6-6a01-4827-ab00-d434456742b7/professionswissen.pdf#page=153 http://psycnet.apa.org/journals/rev/84/3/231/.

Pappas, George Sotiros. 1979. Justification and knowledge : New studies in epistemology. Dordrecht, Holland; Boston: D. Reidel Pub. Co.

Peacocke, Christopher. 1999. Being Known. Oxford; New York: Clarendon Press ; Oxford University Press.

———. 2004. The Realm of Reason. Oxford; New York: Clarendon Press ; Oxford University Press.

———. 2008. Truly Understood. Oxford; New York: Oxford University Press.

———. 2012. Descartes defended, Aristotelian society supplementary volume 86 (1), p. 109-125.

———. 2014. The Mirror of the World : Subjects, Consciousness, and Self-consciousness.

Ross, Lee and Richard E Nisbett. 1991. The Person and the Situation, Perspectives of Social Psychology. Pinter & Martin.

Russell, Bertrand. 1989. Problèmes De Philosophie. Trans. François Rivenc Paris: Payot.

Ryle, Gilbert (2013/1949), The Concept of Mind, Chicago, The university of Chicago Press

Sackur, Jérôme. 2009. L'introspection en psychologie expérimentale. Revue D'histoire Des Sciences.

Schachter, Stanley and Jerome E Singer. 1962. Cognitive, social, and physiological determinants of emotional state. American Psychologist 69 (5): 379-399. http://content.apa.org/journals/amp/20/9/713.

Scharp, Kevin. 2008. Locke's theory of reflection. British Journal for the History of Philosophy 16 (1): 25-63.

Schwitzgebel, Eric. 2003. Self-Ignorance. In Consciousness and the Self. Ed. JeeLoo Liu and John Perry. Cambridge: Cambridge University Press.

———.2008. The unreliability of naive introspection. Philosophical Review 117 (2): 245-273.

———.2011. Perplexities of Consciousness. Cambridge, Mass.: MIT Press.

———.2012. Introspection what ? in Declan Smithies et Daniel Stoljar (dir.), Introspection and Consciousness, Oxford: Oxford University Press.

Shoemaker, Sydney. 1996. The First-person Perspective and Other Essays. Cambridge; New York: Cambridge University Press.

Siewert, Charles P. 1998. The Significance of Consciousness. Princeton, N.J.: Princeton University Press.

———. 2003. Self-knowledge and Rationality: Shoemaker on Self-blindness, in B. Gertler (dir.) Priviledged Access. Aldershot: Ashgate Publishing.

Skinner, B.F. 1938. The Behavior of Organisms: An experimental analysis. Oxford: Appleton-Century.

Sperling, G. 1960. The Information Available in Brief Visual Presentations. Psychological Monographs: General and Applied, 74(11).

Stanovich, Keith E. 1999. Who Is Rational? : Studies of Individual Differences in Reasoning. Mahwah, N.J.: Lawrence Erlbaum Associates.

Titchener, E. B. 1909. Lectures on the Experimental Psychology of the Thought Processes. New York: Macmillan.

Watson, J. B. 1913. Psychology as the behaviorist views it. Psychological Review, 20(2), 158–177.

Wegner, Daniel M. 2002. The Illusion of Conscious Will. Cambridge, Mass.: MIT Press.

Wells, Gary L and Richard E Petty. 1980. The effects of over head movements on persuasion: Compatibility and incompatibility of responses. Basic and Applied Social Psychology 1 (3): 219-230.

Wheatley, Thalia and Jonathan Haidt. 2005. Hypnotic disgust makes moral judgments more severe. Psychological Science 16 (10): 780-784.

Williams, Bernard, (2015/1978), Descartes : the Project of Pure Inquiry, Oxon: Routledge classics.

Wilson, Timothy D. 2002. Strangers to Ourselves: Discovering the Adaptive Unconscious. Harvard: Harvard University Press.

Wittgenstein, Ludwig. 1953. Recherches Philosophiques. Gallimard: Paris (réed. 2004).

———. 1972. Preliminary Studies for the « Philosophical Investigations, » Generally Known As the Blue and Brown Books. Oxford: B. Blackwell.

Wright, Crispin. 1998. Self-knowledge: The wittgensteinian legacy. Royal Institute of Philosophy Supplement 43:101-122.

Pour citer cet article :

Ludwig, P., Michel, M. (2017), « Introspection », version académique, dans M. Kristanek (dir.), l'Encyclopédie philosophique, URL : http://encyclo-philo.fr/introspection-a/