

TRANSPARENCE, CONSCIENCE DE SOI ET SAVOIR-FAIRE

Thibaut Sitterlin
(Nantes Université, CAPHI)

Résumé

Une connaissance de son propre état mental est dite « transparente » quand la capacité du sujet à se l'auto-attribuer ne repose pas sur ses capacités à considérer s'il possède cet état, mais sur ses capacités à considérer le contenu de cet état. Par exemple, ma capacité à m'auto-attribuer la croyance qu'il pleut repose sur mes capacités à considérer s'il pleut. L'objet de cet article est de montrer qu'un modèle représentationnel de la conscience de nos propres états mentaux est incapable de rendre compte de cette transparence. Je défends l'idée d'une conscience de soi intransitive au sens où elle n'est pas représentationnelle et m'efforce de cerner cette notion. J'examine en particulier deux positions qui donnent un sens à cette idée : l'expressivisme inspiré de Wittgenstein, notamment défendu par Vincent Descombes, et le réflexivisme de Matthew Boyle inspiré de Sartre. Les points forts et les difficultés de ces deux approches me conduisent à une conception de la conscience intransitive qui l'apparente à une forme de savoir-faire. Je montre enfin que celle-ci ne se laisse pas définir dans les termes d'une approche intellectualiste.

Abstract

Knowledge of one's own mental state is said to be 'transparent' when the subject's ability to attribute it to himself is not based on his ability to consider whether he possesses this state, but on his ability to consider the content of that state. For example, my capacity to self-attribute the belief that it is raining relies on my capacity to consider if it is raining. The aim of this article is to argue that a representational model of the consciousness of our own mental states is unable to account for this transparency. I defend the idea of an intransitive self-consciousness in the sense that it is not representational and I attempt to clarify this notion in relation to different conceptions that have been proposed. I examine in particular two positions that give meaning to this idea: expressivism inspired by Wittgenstein, notably defended by Vincent Descombes, and Matthew Boyle's reflexivism, inspired by Sartre. The strengths and difficulties of these two approaches lead me to a conception of intransitive consciousness which relates it to a form of know-how. Finally, I argue that it cannot be defined according to an intellectualist approach.

Ἀὐτὸν δὲ νοεῖ ὁ νοῦς κατὰ μετάληψιν τοῦ νοητοῦ.
« L'esprit se comprend par participation à ce qu'il comprend. »
(Aristote, *Métaphysique*, 1072b 19-20)

On parle souvent d'introspection pour désigner à la fois la connaissance qu'un sujet peut acquérir de certains de ses traits de caractère ou dispositions affectives (jalousie, amour, etc.), mais également pour désigner la conscience immédiate de ses propres états mentaux présents (sensations, croyances, intentions, etc.). S'il existe bien un sens neutre du terme « introspection », ce mot véhicule néanmoins l'image d'une sorte de *représentation* de ce qui se passe à l'intérieur de soi. Or cette image est incompatible avec la transparence, c'est-à-dire le fait que, pour certains états mentaux, leur auto-attribution

ne repose pas sur les capacités du sujet à considérer s'il les possède en lui, mais sur ses capacités à considérer directement leur contenu. Par exemple, ma capacité à m'auto-attribuer la croyance qu'il pleut repose sur mes capacités à considérer s'il pleut. L'objet de cette étude est de montrer qu'un modèle représentationaliste de la conscience de nos propres états mentaux est incapable de rendre compte de cette transparence. Je défends l'idée d'une conscience de soi intransitive qui ne se laisse pas analyser en termes de représentation de nos états mentaux. J'expose en premier lieu certains concepts clés (1) et montre ensuite les limites des théories représentationalistes de la conscience de soi (2). Il s'agira alors de proposer une conception adéquate de la notion de conscience intransitive. J'écarterai d'abord les approches phénoménologiques et inférentielles, qui sont à mon sens incapables de rendre compte de la transparence bien comprise (3). J'aborderai alors en détail deux conceptions qui donnent sens à l'idée d'une conscience intransitive non-représentationnelle (4) : l'expressivisme inspiré de Wittgenstein, défendu notamment par Vincent Descombes, et le réflexivisme de Matthew Boyle, inspiré de Sartre. Les points forts et les limites de ces deux approches me conduisent à proposer une conception capacitaire de la conscience intransitive. Cette dernière est pensée comme une forme de savoir-faire, au sens de savoir exercer certaines tendances constitutives d'une attitude mentale selon les contraintes rationnelles appropriées (5). Il apparaîtra enfin que ce savoir-faire ne se laisse pas définir dans les termes d'une approche intellectualiste.

1. Transparence, asymétrie et conscience intransitive

1.1. La transparence

On admet couramment que la conscience que nous avons de nos états mentaux dépendrait d'une représentation de ce qui se passe en nous. On peut néanmoins contester ce présupposé au moins pour certains types d'états mentaux intentionnels, comme les croyances, les perceptions ou les intentions. Si l'on me demande si je crois qu'il va pleuvoir, j'aurais plutôt tendance à regarder ce qui se passe dans le ciel que ce qui se passe en moi. Comme le dit Gareth Evans :

« [...] lorsque nous nous auto-attribuons une croyance, nos yeux sont, pour ainsi dire, ou parfois littéralement, dirigés 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, en lui répondant, exactement sur les mêmes phénomènes extérieurs sur lesquels je porterais mon attention si je répondais à la question « Y aura-t-il une troisième guerre mondiale ? » (Evans 1982, 225)

On parlera de transparence lorsque l'auto-attribution d'un état mental est fondée de cette façon, comme si le sujet regardait « à travers » cet état pour considérer directement son contenu. La transparence peut qualifier soit la connaissance, exprimée dans l'auto-attribution, soit l'état mental lui-même.

(CT) La connaissance en première personne d'un de ses propres états mentaux intentionnels est dite transparente si elle repose, non pas sur la capacité du sujet à considérer s'il possède en lui cet état mental, mais sur sa capacité à considérer le contenu de cet état.

(ET) Un état mental est dit transparent s'il est connu de cette façon.

La transparence est liée aux fameuses phrases paradoxales à la manière de Moore, du type « Je crois que p , mais non- p »¹. En effet si, comme le dit Evans, la transparence est caractérisée par le fait que l'auto-attribution « Je crois que p » est transparente par rapport à la question « Est-ce que p ? », les phrases paradoxales de Moore consistent justement à ne pas la traiter de cette façon.

1.2. L'asymétrie entre la première et la troisième personne du point de vue de la justification

Comme le paradoxe de Moore, la transparence ne concerne que la première personne du présent de l'indicatif. L'attribution d'une croyance à autrui suppose, quant à elle, que l'on considère la psychologie ou le comportement de la personne. Si j'attribue la croyance qu'il pleut à Paul, je dois me fonder sur une observation, une inférence ou un témoignage concernant Paul. Par exemple, je le vois prendre son parapluie et j'en infère qu'il croit qu'il pleut, ou bien il me le dit. On peut me demander « Comment sais-tu que Paul croit qu'il pleut ? ». La question n'a rien de surprenant. En revanche, il serait incongru que l'on me demande « Comment sais-tu que tu crois qu'il pleut ? », comme si l'on s'attendait à ce que je fournisse une preuve concernant ma propre psychologie. Dans le premier cas, la question n'a rien d'étrange, mais l'absence de réponse serait mystérieuse. En revanche, à la première personne, c'est la question qui est étrange, mais l'absence de réponse n'a rien de mystérieux (Campbell 2020). On considère généralement que vous n'avez pas à justifier votre auto-attribution quand il s'agit de votre propre état mental présent – à la différence d'une auto-attribution d'un état passé. Ou bien, s'il s'agit d'états transparents, leurs auto-attributions, lorsqu'elles sont justifiées, le sont directement au regard de leur contenu. Il y a donc une asymétrie entre la première et la troisième personne du point de vue de la justification pour ce qui concerne les attributions d'état mentaux, asymétrie en vertu de laquelle l'auto-attribution bénéficie généralement d'un privilège ou d'une autorité en ceci que la sincérité de l'assertion implique sa vérité. Si vous dites sincèrement que vous croyez qu'il pleut, c'est que vous croyez qu'il pleut.

1.3. Conscience transitive et conscience intransitive

Il se peut néanmoins qu'à l'égard de certains états mentaux nous ne bénéficions pas d'une telle autorité ou que leur auto-attribution nécessite une prise en compte de notre propre psychologie. C'est le cas, par exemple, des dispositions de notre caractère, comme la jalousie, l'irritabilité, etc. mais aussi des états mentaux inconscients que théorisent autant la psychanalyse que les sciences cognitives. Peut-être avons-nous des croyances inconscientes. Sans doute avons-nous aussi des perceptions inconscientes. On pensera par exemple au phénomène de la vision aveugle qui consiste, pour certains individus ayant leurs aires corticales visuelles endommagées, à pouvoir percevoir des objets, puisqu'ils y réagissent, mais sans en avoir d'image visuelle consciente. Dans ces exemples, il s'agit d'états doués de contenu intentionnel pour lesquels, néanmoins, le principe de transparence ne fonctionne pas et pour lesquels la question « Comment savez-vous que

¹ Le paradoxe vient du fait qu'en disant « Je crois que p » je suis engagé vis-à-vis de la vérité de p , ce qui n'est pas le cas en disant « Il croit que p » ou « Je croyais que p ».

vous vous êtes dans cet état ? » a un sens. Il n'est donc pas exclu que nous puissions *prendre conscience* de ces états, par observation, inférence ou témoignage. Par exemple, je peux prendre conscience de ma croyance inconsciente que personne de m'aimera jamais et me l'auto-attribuer (Finkelstein 2003). Une personne sujette à la vision aveugle peut savoir que son système visuel représente les objets en face d'elle. Il y a donc un sens à dire qu'elle est consciente de sa perception sans que cette perception elle-même soit consciente. Il en va peut-être de même pour ma croyance inconsciente. Il se peut que j'en prenne conscience, mais sans l'assumer en première personne comme quelque chose que je juge crédible. Mon auto-attribution ne sera pas, dans ces cas, transparente, mais dépendra de la considération de certains éléments de ma propre psychologie au même titre que l'attribution d'un état mental à autrui. On dira dans ce type de cas que mon rapport à ma croyance est aliéné ou de spectateur (Moran 2001), parce qu'il n'est pas authentiquement en première personne, mais analogue à une attribution en troisième personne.

On peut donc, en un sens, être conscient d'un de ses états mentaux sans que cet état mental lui-même soit *conscient* en un autre sens. On peut être conscient *de* croire ou percevoir sans croire ou percevoir *consciemment*. Cette idée peut être illustrée à l'aide du paradoxe de Moore. Si les phrases à la manière de Moore sont généralement paradoxales, il y a peut-être des cas où le paradoxe s'efface. Par exemple si, en disant « Je crois que *p*, mais non-*p* », je veux dire quelque chose comme « Je crois *inconsciemment* que *p*, mais non-*p* ». Le paradoxe disparaît précisément parce que dans la première partie de la phrase, je n'envisage pas la question de savoir si je crois que *p* comme transparente par rapport à la question portant sur *p*. Dans un cas, j'exprime un point de vue *sur* ma croyance, dans l'autre j'exprime ma croyance elle-même d'une façon telle que je manifeste par là que j'adhère consciemment à son contenu. On distinguera en ce sens la *conscience transitive* (être conscient de sa croyance ou de sa perception que *p*) et la *conscience intransitive* (avoir une croyance ou une perception *consciente* au sens de croire ou percevoir *consciemment* que *p*). Dans la mesure où il semble difficile de considérer qu'un état mental puisse être conscient indépendamment d'un rapport à un sujet, je considérerai que la locution adjectivale équivaut à la locution adverbiale.

On pourrait estimer que l'hypothèse des croyances inconscientes est trop polémique pour être acceptée telle quelle. Notons néanmoins que si l'hypothèse d'états mentaux inconscients est commode pour expliquer la distinction entre les deux types de conscience, elle n'est pas décisive pour en justifier la pertinence. Peut-être n'existe-t-il rien de tel, en réalité, que des croyances inconscientes. Mais si l'auto-attribution d'une telle croyance est illusoire, elle est néanmoins possible et demandera des justifications épistémiques pour être rationnelle. En outre, on peut tout à fait adopter un point de vue spectral sur une croyance qui par ailleurs est consciente. Quelqu'un qui croit en Dieu pourrait très bien objectiver sa croyance en s'interrogeant sur les causes – plutôt que les justifications – qui expliquent son état, en se demandant, par exemple, si sa croyance est seulement le fruit de son éducation. Il n'envisage plus ici sa croyance au regard de son contenu comme quelque chose qu'il assume en première personne, mais comme un état psychologique dont il s'efforce d'expliquer la genèse en lui. Il ne considère alors pas tant *ce* qu'il croit pour lui-même, que *le fait* de le croire. Ce faisant, il adopte à l'égard de sa croyance un point de vue de spectateur, au sens où il l'envisage comme de l'extérieur, en troisième personne.

2. Critiques des analyses représentationalistes de la conscience de soi

2.1. Transparence et représentation

La conscience transitive est généralement analysée selon un schéma intentionnel de type représentationnel. Être conscient de *p* implique de se représenter *p*. Plus précisément, c'est se représenter un contenu selon une attitude ou un mode psychologique particuliers. On peut croire qu'il pleut, le craindre, le désirer, etc. Le contenu lui-même ne change pas, mais il est pensé selon différentes attitudes intentionnelles qui sont autant d'états mentaux ayant la propriété d'être à propos de quelque chose dont le sujet est conscient parce qu'il se le représente. On pourrait alors penser que ce qui rend un état mental conscient (intransitivement), c'est le fait que le sujet *en* est conscient (transitivement) d'une façon spécifique. Si, comme on vient de le voir, la conscience transitive n'est pas suffisante pour qu'un état soit conscient, elle en est peut-être cependant une condition nécessaire.

Mais le principe de transparence suggère justement que ce schéma ne fonctionne pas. En effet, si l'on en croit Evans, ce qui caractérise la transparence, c'est que l'auto-attribution de ma croyance qu'il pleut repose sur ma capacité de répondre à la question de savoir s'il pleut. Il semble donc qu'en jugeant

(1) Il pleut

je sois autorisé à penser immédiatement

(2) Je crois qu'il pleut

Je n'ai besoin d'aucune information supplémentaire pour passer de (1) à (2). Pourtant, le *contenu* de ma première pensée ne fait aucunement référence à ma croyance. Il pourrait pleuvoir sans que je ne le croie ; et en règle générale, la vérité de *p* ne fournit pas de critère pour dire que je crois que *p*. Néanmoins, je suis autorisé à passer immédiatement de (1) à (2) et (2) exprime bien une forme de conscience de ma croyance. En conséquence, il semble qu'il doive y avoir une forme de conscience relative à sa propre croyance, qui est impliquée dans le fait de penser (1) et qui ne suppose pas la représentation de cette croyance (Boyle 2024). Nous sommes donc conduits à l'hypothèse d'une conscience, intransitive, qui ne se laisse pas expliquer en termes de représentation d'un contenu.

2.2. Critique des théories des représentations d'ordre supérieur

Jean-Paul Sartre (1943) soutenait une thèse de ce type en disant que toute conscience positionnelle suppose une conscience de soi non-positionnelle. Il s'efforçait de le justifier en réduisant à l'absurde la thèse opposée selon laquelle tout état conscient dépend d'une représentation d'ordre supérieur. Mais il est passé bien trop rapidement sur cette critique, et c'est pourquoi il convient de la compléter.

Selon les théories des représentations d'ordre supérieur, un état conscient est simplement un état *dont* le sujet est conscient et il en est conscient en raison d'une représentation de second ordre (Rosenthal 1986). Ma croyance que *p* est consciente en vertu d'un état mental d'ordre supérieur qui a pour contenu « je crois que *p* ». Pour que ma croyance soit consciente, il faut que je me représente que je le crois. Si je n'ai pas de représentation (perception ou pensée) de second ordre, je ne peux pas croire *consciemment* ou, ce qui revient au même, ma croyance ne peut pas être *consciente*. Bien sûr, il faut rendre compte de la différence avec les cas où la relation à ma croyance est aliénée. Il suffit alors de

postuler que, dans les cas normaux, il existe un lien plus direct entre ma représentation d'ordre supérieur et ma croyance de premier ordre. On pourra dire, par exemple, que l'état de premier ordre cause directement le second état ou qu'il n'existe pas d'observation ou d'inférence conscientes entre les deux.

Mais il y a une première difficulté avec cette hypothèse : pour être conscient de quelque chose, il faut être dans une attitude consciente à l'égard de cette chose (cf. Rowlands 2001). J'appellerai « principe de conscienciosité » l'énoncé suivant :

(PC) A est conscient que p (ou de O) ssi A se représente consciemment que p (ou se représente consciemment O).

Si je suis conscient *de la pluie*, c'est parce que je pense ou perçois *consciemment* la pluie. Cette supposition est nécessaire si l'on veut distinguer une représentation consciente et une représentation inconsciente (par exemple, la vision aveugle), comme le souhaite le théoricien des représentations d'ordre supérieur lui-même. Mais si ce qui fait qu'un état mental est conscient est le fait qu'il est représenté dans un état d'ordre supérieur, alors on est conduit dans une régression à l'infini. Pour que je sois conscient de ma croyance ou de ma perception qu'il pleut, il faut que je me la représente dans un état d'ordre supérieur qui à son tour devra être conscient, etc.² Comme le dit Sartre (1943, 18-19), soit on postule une conscience inconsciente, ce qui est une absurdité, soit on est pris dans une régression. Certains répondront qu'il n'est pas besoin d'une représentation de second ordre occurrente, mais qu'il suffit d'être *disposé* à se représenter l'état de premier ordre (Carruthers 1996 ; 1998). On éviterait ainsi une régression à l'infini en acte. Mais il y a un autre problème. Selon les théories des représentations d'ordre supérieur, le fait qu'un état soit conscient est constitué par l'auto-attribution actuelle ou potentielle de cet état. C'est donc l'auto-attribution qui fonde la conscience et non l'inverse, ce qui peut déjà paraître contre-intuitif, puisque l'on aurait tendance à penser, au contraire, que c'est la conscience de mon état qui fonde son auto-attribution (O'Brien 2007, 97). Mais surtout, si l'on reconnaît, d'après le principe de conscienciosité, que l'on ne peut être conscient de quelque chose que si on se le représente consciemment, alors la théorie des représentations d'ordre supérieur devrait impliquer que, puisqu'un état mental n'est conscient qu'en vertu de son auto-attribution effective ou potentielle, notre conscience des choses et états de choses *du monde* dépend de l'auto-attribution de l'état qui le représente. En conséquence, si je vois qu'il pleut, ce qui me rend conscient *de la pluie*, c'est que je puisse m'auto-attribuer une perception de la pluie. Et la possibilité de l'auto-attribution doit précéder celle de la pensée consciente sur la pluie. Il y a alors deux conséquences.

D'abord, contrairement à ce que révèle l'analyse de la transparence, l'auto-attribution de la croyance que p ne peut pas être justifiée par la considération de p et des raisons de croire que p , puisque la conscience de p elle-même dépend de l'auto-attribution de cette croyance. Cette auto-attribution serait donc indépendante de la considération de son contenu et des justifications pour ce contenu. Si l'on compare avec l'analyse précédente de la transparence, on doit dire que le sujet qui connaît de cette façon ses états mentaux est davantage *incité* en vertu d'un mécanisme de représentation à penser qu'il les possède que *justifié* au regard de leur contenu. En conséquence, une représentation d'ordre

² Cette objection s'applique autant à une théorie des pensées d'ordre supérieur qu'à une théorie des perceptions d'ordre supérieur.

supérieur n'est jamais qu'une sorte de rapport sur son état mental, non une expression directe de celui-ci.

Ensuite, le principe de conscosité devrait conduire un partisan des représentations d'ordre supérieur à juger incohérente l'attribution d'une conscience, non seulement de leurs propres états mentaux, mais également *de leur environnement* à des êtres qui ne disposent pas des capacités conceptuelles pour s'auto-attribuer des jugements, des croyances ou des perceptions. Un nouveau-né ne pourrait pas être conscient, par exemple, de la présence de sa mère ou un chat de la présence d'une souris en face de lui, puisque, étant incapables de se l'attribuer, ils ne pourraient en avoir de perception consciente. Et cela vaut d'ailleurs tout autant pour les sensations et les émotions. Mais après tout, peut-être les nouveau-nés et les chats sont-ils tous des voyants aveugles ou des animaux-machines ? La question qui se pose est de savoir si c'est là quelque chose que l'on serait prêt à accepter en dehors des conséquences de la théorie.

2.3. La thèse de l'auto-représentation

Une autre option de l'approche représentationnelle de la conscience de soi, consiste à dire que les états conscients auraient la particularité de se représenter eux-mêmes (Brentano 1944 ; Kriegel 2009). Une pensée comme « Il pleut » aurait en fait deux contenus : le contenu primaire « Il pleut » et le contenu secondaire, implicite, « Je pense par la présente qu'il pleut » ou quelque chose de tel. L'intérêt d'une telle approche par rapport à la précédente, c'est qu'elle fait du caractère conscient de l'état mental une propriété inhérente au fait d'être dans cet état.

Il convient cependant de distinguer deux significations du pronom dans l'expression « se représenter ». Soit il désigne l'attributaire, celui à qui quelque chose est représenté, soit il désigne ce qui est représenté, dans un usage réflexif. Faire figurer le sujet dans le contenu, satisfait précisément ce second usage. Or « représenter » est justement un verbe *transitif*. L'expression « se représenter » dans son usage réflexif, indique que l'agent et le patient sont le même, comme dans « se soigner », Dès lors, se représenter soi-même, c'est nécessairement, comme pour le médecin d'Aristote qui se soigne lui-même, se représenter *en tant qu'autre*. Si un accusé se représente lui-même en assurant sa propre défense, ce n'est pas en tant que défenseur qu'il est représenté, mais en tant qu'accusé. En raison de cette dualité, les mêmes difficultés qui mettent à mal une théorie de la représentation d'ordre supérieur menacent également une théorie de l'auto-représentation. Par ailleurs, cette doctrine postule qu'un état mental conscient est un état qui se représente lui-même ainsi que le sujet de cet état (« Je pense par la présente que *p* »). Mais en quoi cela suffirait-il à rendre conscient un état mental ? Après tout, il n'est pas exclu a priori que l'on puisse avoir des pensées inconscientes et auto-référentielles sur soi-même. « Je suis une phrase en français » est une phrase qui se représente elle-même. Sans doute mime-t-elle la conscience de soi, mais elle n'est pas consciente. L'auto-représentation ne suffit donc pas à la conscience. Il faudrait donc encore qu'un état mental se représente *consciemment* lui-même pour être conscient. Et le risque de la régression ou de la circularité apparaît : pour qu'un état soit conscient, il faut qu'il se représente consciemment lui-même ; et pour cela, il devra se représenter lui-même se représentant, etc.

Arrivé à cette étape, il semble bien que l'on doive postuler une forme de conscience intransitive qui n'est pas de l'ordre de la représentation d'un contenu, mais que suppose toute représentation consciente. Le problème devient alors de savoir en quoi elle consiste.

3. Conscience intransitive et phénoménalité

3.1. La thèse phénoménologique

Un certain nombre de philosophes qui admettent l'idée d'une conscience intransitive en font une caractéristique de l'expérience (Henry 1963 ; Zahavi 1999). Elle ne devrait donc concerner que des états doués de propriétés phénoménales, comme les sensations ou les émotions. Certains états mentaux sont motivants par eux-mêmes et ont la raison de leur auto-attribution directement dans leur occurrence plutôt que dans une représentation de cette occurrence. Quand je ressens une démangeaison, celle-ci peut être un motif pour me gratter sans qu'il soit besoin que j'en possède un état de conscience distinct (Peacocke, 1998). De façon analogue, on peut supposer que certains états mentaux doués de phénoménalité sont directement la raison de leur auto-attribution sans que celle-ci n'en suppose de représentation préalable. Mais ce modèle est-il exportable pour des états mentaux intentionnels et dispositionnels, comme les croyances ou les intentions ? Y a-t-il une phénoménologie distinctive de la croyance ? On peut en douter si on la conçoit comme un état dispositionnel plutôt qu'il y ait bien quelque chose de tel, fonder la conscience de sa propre croyance sur une phénoménologie distinctive ne nous fait pas réellement sortir d'une relation spectatorielle à notre croyance et cela met à mal l'idée de transparence. A supposer qu'il y ait un « effet que cela fait » de croire, la croyance, à la différence d'une sensation comme la douleur, est difficilement réductible à cette seule phénoménalité. Il faudrait encore fonder l'auto-attribution de la croyance à partir de l'expérience caractéristique qui l'accompagne. Il s'agirait alors de dire quelque chose comme « Je sais que je crois que *p*, parce que je sens que je le crois ». Mais cela est indépendant de la considération de *p* et des justifications possibles pour *p*. Pour cette raison, une telle auto-attribution demeure aliénée. On pourrait vouloir s'en sortir en attribuant directement une qualité phénoménale au contenu, en référence à la vénérable notion d'évidence. Mais on ne saurait raisonnablement supposer que toutes nos croyances reposent sur l'évidence apparente de leur contenu.

3.2. Le modèle inférentialiste

Une solution envisageable serait de dire que l'auto-attribution de croyance est en réalité *inférée* à partir d'un état phénoménalement conscient distinct. C'est ce que fait, par exemple, Christopher Peacocke (1998). Selon lui, une telle auto-attribution repose sur un jugement qui est un acte mental de ce type. La conscience immédiate de mon jugement que *p* me justifiera à m'auto-attribuer une croyance que *p* dans la mesure où, d'après Peacocke, le jugement est, à l'instar du souvenir, conceptuellement lié à la croyance. De même que se souvenir que *p* suppose que l'on croie que *p*, de même juger que *p* exprime généralement ou produit la croyance que *p*. C'est ainsi que Peacocke explique que, conformément au principe de transparence, je puisse passer sans information supplémentaire de « Il pleut » à « Je crois qu'il pleut ».

Mais là encore, on n'a affaire qu'à un rapport aliéné à sa propre croyance (Boyle 2024, chap. 2.4³). Par une telle inférence du jugement à la croyance, je suis seulement *incité* à

³ Boyle applique le même type de critique à l'inférentialisme de Byrne (2018).

penser que je crois qu'il pleut, non pas justifié à partir du contenu de mon jugement lui-même. Au fond, ce qui me porte à m'auto-attribuer une croyance, c'est l'idée que si je juge qu'il pleut, il faut bien que je le croie. Je *découvre* seulement que je crois, comme je pourrais le découvrir de quelqu'un d'autre. Par ailleurs, pour que je sois autorisé à passer directement du jugement sur la pluie au jugement sur ma propre croyance, il faudrait que je sache déjà que mon jugement est l'expression d'une croyance et non pas, par exemple, d'une simple assomption.

On en arrive donc à l'idée qu'il doit y avoir une forme de conscience qui n'est ni représentationnelle, ni fondée sur une inférence, ni réductible à une conscience phénoménale. Cette forme de conscience intransitive est celle qui est supposée dans l'idée de transparence. Je voudrais alors examiner deux approches qui s'appuient sur l'hypothèse d'une telle conscience.

4. Les thèses de l'équivalence

Il y a une façon simple de rendre compte du fait que l'on est autorisé à passer immédiatement et sans inférence de

(1) Il pleut

à

(2) Je crois qu'il pleut

Il suffit de soutenir que les deux énoncés disent d'une certaine façon la même chose. C'est d'ailleurs ce que suggère les phrases paradoxales de Moore. Si la phrase « Je crois qu'il pleut, mais il ne pleut pas » est paradoxale, c'est parce que, à tout le moins dans certains usages courants des expressions qui la composent, elle est contradictoire. C'est donc que les deux parties de la phrase disent la même chose. Puisqu'il se peut qu'il pleuve sans que je le croie, comme on l'a déjà dit (2. 1), il semble exclu que l'énoncé (1) comprenne dans son contenu une référence à ma croyance – et la critique du modèle de l'auto-représentation rend peu pertinente l'idée d'un contenu secondaire implicite. Il reste alors deux possibilités. Si (1) ne parle pas de moi, on pourrait considérer qu'en réalité (1) et (2) parlent *tous deux* de la pluie. Ils réfèrent à un fait ou une vérité du monde et le « je crois que » de l'énoncé (2) ne change pas la référence de (1). L'énoncé (2) ne décrit pas un état mental du sujet, mais tout comme (1), il *exprime* directement la croyance du sujet concernant la pluie. C'est la position expressiviste associée à Wittgenstein et magistralement défendue par Descombes (2014, chap. 7).

Selon une autre approche, (1) et (2) disent tous deux, *d'une certaine façon*, quelque chose de moi. (2) est une explicitation de quelque chose qui serait implicite dans l'assertion de (1) et que le sujet capable de *réflexion* sur ses propres états de croyance serait en mesure d'explicitier. Selon cette interprétation, (2) fait bien référence à soi et cette référence serait implicitement comprise, non dans le contenu de (1), mais dans la manière de le penser. Dans ce cas, l'équivalence n'est pas directement entre (2) et (1), mais entre (2) et un présupposé implicite de (1). C'est la position réflexiviste défendue par Matthew Boyle (2024) qui revendique une conception sartrienne de la conscience intransitive comme conscience tacite et préréflexive.

4.1. L'expressivisme

A vrai dire, il existe deux grandes variantes de l'expressivisme : la version grammaticale, disons de « l'ancienne école », notamment défendue par Descombes, qui refuse de traiter les expressions comme « Je crois que *p* » ou « J'ai mal » comme des auto-attributions ; et le néo-expressivisme (Finkelstein 2003 ; Bar-On 2004 et, dans une certaine mesure, Campbell 2020), qui l'accepte. La différence de ces deux versions est, me semble-t-il, très instructive. Je commence par la plus ancienne.

Descombes présente la thèse expressiviste au sujet de la croyance sous l'appellation de « solution grammaticale » au paradoxe de Moore. L'appellation vient de la distinction que faisait Wittgenstein entre la grammaire profonde et la grammaire de surface de nos énoncés. La grammaire de surface d'une phrase comme « Il pleut » est identique à celle d'une phrase comme « Il boude ». Pourtant, du point de vue de la grammaire profonde, si « Il boude » est une authentique attribution, « Il pleut » ne l'est pas, car ici le pronom « il » ne renvoie à aucun sujet auquel on pourrait attribuer l'action de pleuvoir. De façon analogue, un expressiviste de l'ancienne école dirait que si « Je crois qu'il pleut » ressemble à une attribution comparable à « Il croit qu'il pleut » du point de vue de la grammaire de surface, il n'en est rien au regard de la grammaire profonde, sinon les phrases à la manière de Moore n'auraient rien de paradoxal. L'énoncé (1) porterait sur moi et l'énoncé (2) sur le monde.

En vérité, dit l'expressiviste, en disant « Je crois qu'il pleut », *j'exprime* ma croyance en décrivant *ce que* je crois. Je ne *décris* pas *le fait que* je crois. Il en va exactement comme lorsque je dis simplement « Il pleut ». J'exprime alors ma croyance en parlant de la pluie. En disant « Je crois que... » je n'identifie pas un sujet qui serait moi-même pour lui attribuer une croyance, comme je le fais au contraire dans des attributions de croyance en troisième personne.

Selon l'expressivisme, l'énoncé « Je crois que *p* » exprime une croyance un peu comme un sourire exprime la joie. En le prononçant, je parle *à partir de* ma croyance et non *à son propos*. Je « laisse libre cours » à ma croyance, un peu comme la grimace, ou « Aïe » laissent libre cours à la douleur. Une expression verbale de croyance est en ce sens une sorte d'« aveu » (*avowal*⁴) qui n'est pas plus une description de son propre état mental que ne le sont un sourire, une grimace ou « Aïe ». Il en va de même pour « J'ai mal », qui est une sorte de raffinement linguistique du cri de douleur. Le sujet ne fait pas un compte rendu relatif à une information qu'il aurait enregistrée sur une base empirique portant sur lui-même, comme une perception, une observation ou une inférence. Certes, parce que leur grammaire est propositionnelle, les aveux sont porteurs de valeur de vérité, à la différence d'un sourire. Mais cette valeur de vérité est évaluable seulement du point de vue de la sincérité, tout comme un sourire peut être sincère ou non. L'aveu n'a, en effet, pas davantage de fondement épistémique que le sourire.

Sur ce dernier aspect de l'expressivisme, les anciens et les modernes sont parfaitement d'accord. La différence est que les anciens, contrairement aux nouveaux, refusent de comprendre les aveux comme des véritables auto-attributions et, en conséquence, comme des expressions d'une forme de connaissance de soi. Un argument parfois attribué à Wittgenstein consiste à dire qu'il n'y a d'intérêt à parler de connaissance que là où il y a

⁴ La traduction par « aveu » est certainement contestable. Mais étant donné l'usage assez libéral de l'expression « J'avoue » par la jeunesse actuelle, peut-être serai-je pardonné si je force un peu le sens premier du mot. Entendons ici, une expression linguistique d'un état mental conscient comportant une forme au moins apparente d'auto-attribution.

une possibilité d'erreur, d'ignorance ou de découverte, ce qui n'est pas le cas des états mentaux pour lesquels nous jouissons d'une autorité en première personne, au sens où la sincérité implique la vérité. Pour le grammairien, les expressions d'états mentaux ne sont donc pas des expressions de la connaissance de ces états, du moins au sens de ce que j'ai appelé une conscience transitive⁵. Mais cela n'empêche pas de considérer que l'énoncé « Je crois que *p* » est bien l'expression d'une croyance intransitivement consciente, ne serait-ce que parce que le simple énoncé « *p* » auquel il est équivalent exprime généralement déjà une telle croyance consciente. Cette conscience n'aurait cependant pas de valeur épistémique.

Cette version non-épistémique de l'expressivisme est souvent rejetée au motif qu'elle devrait impliquer une conséquence absurde : il y aurait une seule personne au monde qui serait incapable de connaître mon état mental présent et de me l'attribuer, à savoir moi-même (Moran 2001). Cette thèse rendrait également inconcevables des inférences du type « Je crois que *p* ; Paul croit que *p* ; donc nous sommes deux à le croire », une telle inférence supposant de traiter la première prémisse comme une attribution. Enfin, cette thèse ne tiendrait pas suffisamment compte du fait que l'on peut également découvrir, ignorer ou mal identifier certaines de ses propres croyances actuelles. A cela, le grammairien répond qu'il existe bien un usage secondaire de l'expression « Je crois que » qui relève d'une sorte d'auto-attribution. Mais cet usage est aliéné, comme lorsque l'on se situe du point de vue d'un interlocuteur, ou bien lorsque l'on parle de ses croyances inconscientes, ou encore dans certains cas que les psychologues qualifient de dépersonnalisation où un sujet devient comme spectateur de lui-même⁶.

Supposons, néanmoins, une autre situation simple. Quelqu'un demande à une assemblée « Qui ici croit que *p* ? ». Vous levez la main et on s'enquiert ensuite de savoir pourquoi vous avez levé la main. Vous répondez simplement « Parce que je crois que *p* ». Dans une telle situation, vous parlez bien de vous-même, du *fait* que vous croyez. Il s'agit d'une auto-attribution⁷. Le grammairien pourrait alors dire que le fait même de s'attribuer un état mental montrerait que le sujet prend à l'égard de celui-ci une position de spectateur, par exemple dans un contexte d'interlocution où il se situe du point de vue d'autrui. L'auto-attribution serait un usage particulier et secondaire des expressions comme « Je crois que », tout comme il existe un usage primaire, factif, des verbes de perception et un usage secondaire par lequel nous décrivons nos impressions. C'est peut-être vrai. Mais la question est alors de savoir comment il se fait que l'on puisse passer *immédiatement* d'une assertion qui porte sur le *contenu* de la croyance à une assertion qui porte sur le *fait* de croire. Dans notre exemple, je n'ai besoin d'aucune information supplémentaire pour m'auto-attribuer une croyance en répondant à la question « Qui croit que *p* ? ». Je peux passer directement de « *p* » ou « Je crois que *p* », au sens d'une expression de croyance équivalente à « *p* », à « Je crois que *p* » au sens d'une auto-attribution. Comment donc ce passage est-il possible, si les deux énoncés ne parlent pas de la même chose, puisque le

⁵ Voir la distinction que fait Descombes entre conscience et connaissance de soi dans Descombes & Larmore 2009, 183.

⁶ Comme le soldat en état de sidération sur le champ de bataille que décrit Descombes (2014, 336-337). Il sait que l'ennemi approche mais est tétanisé au point qu'il ne réagit pas en fuyant. Il se comporte comme s'il disait « L'ennemi arrive et je ne le crois pas ».

⁷ Cet exemple se trouve chez Hacker 2007, lui-même expressiviste ancien. Mais il l'exploite dans un autre contexte.

premier parle du monde et le second de soi ? Dans notre exemple, celui qui s'auto-attribue une croyance n'est pas non plus aliéné par rapport à celle-ci. Il n'en va pas comme s'il se disait « Il me semble croire que p ». L'auto-attribution en première personne n'est pas un jugement aliéné dans le même sens que « Ludwig Wittgenstein croit que p » dit par Ludwig Wittgenstein ou « Je crois inconsciemment que p » peuvent l'être. Il n'est pas davantage question de dépersonnalisation psychologique. Qu'est-ce alors que cette auto-attribution, si elle n'est pas non plus une authentique expression de croyance équivalente à l'assertion « p » ?

En vérité, on peut supposer qu'il n'y a pas seulement deux usages de « Je crois que p », mais bien trois :

(i) « Je crois que p » est une expression de croyance équivalente à « p ».

(ii) « Je crois que p » est une auto-attribution aliénée en ce sens que le sujet envisage sa croyance comme en troisième personne. Il pourrait dire « Il me semble croire que p » ou « Je crois faussement que p » comme il le dirait de quelqu'un d'autre.

(iii) « Je crois que p » est une auto-attribution non-aliénée, comme dans le cas de la réponse à la question « Qui ici croit que p ? ».

L'expressivisme grammatical ne donne pas vraiment de sens à (iii) – Descombes n'en parle d'ailleurs même pas. En outre, en se focalisant sur (i), le grammairien donne au verbe « croire » dans « Je crois que » un sens factif ou quasi-factif. Dire « Je crois que p , mais non- p » est paradoxal comme « Je sais que p , mais non- p » l'est – avec cette différence que pour le verbe « savoir » la contradiction fonctionne quels que soient le temps et la personne grammaticale. Selon l'approche grammaticale, « Je crois que... » suppose, en effet, la vérité du contenu qu'il introduit. C'est du moins ce que suggère Wittgenstein, lorsqu'il écrit que « s'il existait un verbe signifiant "croire faussement" il n'aurait pas de sens à la première personne du présent de l'indicatif » (Wittgenstein 2004, 270). Le problème, c'est que l'usage factif des expressions d'états mentaux ne bénéficie pas de l'*autorité* en première personne. Je peux dire sincèrement que je sais qu'il pleut ou que je vois un arbre au sens factif du verbe « voir », sans que cela n'implique que mes assertions soient vraies. Mais il n'en va pas de même avec « Je crois que p ». Même si p est faux, il demeure vrai que si je dis sincèrement que je le crois, alors c'est que je le crois. Et ceci n'a de sens que si l'on reconnaît une forme d'auto-attribution non-aliénée (iii). L'expressivisme grammatical donne bien une explication au principe de transparence, mais néglige finalement l'autorité de la première personne pour les états comme la croyance⁸.

C'est ici qu'intervient le néo-expressivisme. Comme son illustre prédécesseur, il reconnaît la spécificité expressive des « aveux ». Mais contrairement à celui-ci, il considère ces

⁸ On verra néanmoins plus bas que ce point n'est peut-être pas toujours problématique pour certains verbes psychologiques par rapport auxquels la transparence n'implique pas l'autorité (5.1). Quoi qu'il en soit, le cas de la perception n'est pas de même nature que celui de la croyance. Avec les verbes de perception, il y a deux usages bien distincts, puisque je peux dire aussi bien « Je vois un arbre, mais il n'y en a pas » (sens non-factif) que « Je ne vois pas d'arbre (j'ai seulement l'impression d'en voir un), car il n'y en a pas » (sens factif). Mais on n'a pas cette distinction avec « Je crois que ». Les usages (iii) et (i) de « Je crois que » sont corrélés de façon plus étroite que les usages factifs et non-factifs des verbes de perception.

expressions comme d'authentiques auto-attributions. Le paradigme implicite est alors l'auto-attribution non-aliénée (iii). Il pourrait donc être légitime de dire que le sujet qui s'auto-attribue ainsi un état mental exprime bien une forme de connaissance de soi, même si celle-ci n'est pas fondée sur autre chose que sur le fait même d'être dans l'état auto-attribué. La différence avec les attributions en troisième personne, c'est qu'il n'y a pas de base de justification épistémique autre que l'état exprimé lui-même. Dorit Bar-On (2004) soutient en ce sens que la connaissance est seulement « collatérale » et non pas préalable à l'aveu. Alors qu'en disant « J'ai mal » ou « Je crois qu'il pleut » j'exprime immédiatement ma douleur ou ma croyance, en attribuant ces états mentaux en troisième personne ou de façon aliénée, je ne fais qu'exprimer un point de vue sur l'état mental.

Quelles sont alors les conséquences pour la notion de conscience intransitive ? Les expressivistes adhèrent souvent à une conception capacitaire de la connaissance, héritée de Wittgenstein et de Ryle. Savoir, c'est pouvoir faire un certain nombre de choses. De la même façon, la conscience intransitive pourrait être définie en termes de capacités et il n'y aurait pas de raison, selon les néo-expressivistes, de refuser de lui accorder une valeur épistémique, tout en lui reconnaissant une irréductible spécificité en raison de l'absence de justification des aveux. Ainsi, la conscience intransitive serait une *capacité à exprimer son état mental dans une auto-attribution non-aliénée*. Si en disant « Je crois que *p* » j'exprime bien ma croyance et non pas seulement un point de vue sur elle, alors mon auto-attribution est l'exercice d'une conscience intransitive et l'état exprimé ou « avoué » est un état conscient, pas simplement un état *dont* je suis conscient au sens d'une conscience transitive. C'est la différence avec les auto-attributions aliénées. Comme l'écrit David Finkelstein :

« L'état mental d'une personne est conscient si elle a la capacité de l'exprimer simplement en l'auto-attribuant. S'il lui manque une telle capacité par rapport à l'un de ses états mentaux, celui-ci est inconscient. » (Finkelstein 2003, 120 – en italiques dans l'original)

Plus récemment, Lucy Campbell (2020) a proposé une définition de la connaissance de soi dans le cadre plus large d'une conception de la connaissance propositionnelle comme capacité de réponse rationnelle, selon laquelle connaître, c'est être capable de répondre rationnellement à un fait. Par exemple, ma connaissance du fait qu'il pleut est constituée d'un ensemble de capacités à faire ou éprouver certaines choses *en raison de* ce fait, comme prendre mon parapluie, me réjouir de la fin de la sécheresse, etc. La spécificité de la connaissance de soi en première personne pourrait alors résider dans la capacité d'exprimer ou d'« avouer » de façon rationnelle le fait d'être dans un état mental, comme lorsqu'une personne crie d'une façon particulière en direction d'autrui pour lui communiquer sa douleur, ou lorsqu'elle s'auto-attribue une croyance comme dans (iii). On préserve une forme d'expressivisme, mais avec une conception épistémique décomplexée.

Cette redéfinition de l'expressivisme souffre cependant, par rapport à l'ancienne version, d'un défaut majeur pour ce qui concerne l'expression des états intentionnels et transparents. En se focalisant sur (iii), elle donne certes une place de choix à l'autorité de la première personne, mais au détriment de la transparence⁹. Le néo-expressivisme a

⁹ C'est flagrant chez Finkelstein (2003, Post-scriptum) et Bar-On (2004, chap. 4), qui s'efforcent de relativiser l'intérêt de cette notion.

tendance à traiter de façon similaire l'expression d'une douleur et l'expression d'une croyance, accordant finalement peu d'importance à l'usage (i) et à la spécificité d'états intentionnels qui peuvent être à la fois dispositionnels plutôt qu'occurrents et sans phénoménalité distinctive.

Qui plus est, il est très contestable de soutenir que les auto-attributions pour lesquelles nous jouissons de l'autorité de la première personne ne reposent pas sur des fondements épistémiques autres que le fait d'être dans l'état auto-attribué. Nombre de nos croyances ou de nos intentions sont bel et bien justifiées – et très souvent sur des bases empiriques – au regard de leur *contenu*. Ce n'est certainement pas un phénomène marginal pour comprendre la spécificité de leur caractère conscient, même si, bien sûr, toutes nos croyances ne sont pas justifiées. On le voit donc, une difficulté majeure pour l'expressivisme dans ses différentes variantes, anciennes et nouvelles, est de rendre compte de l'articulation des usages (i) et (iii).

Notons enfin que la conception expressiviste de la conscience de soi intransitive pose également problème. D'abord, dans la mesure où elle est largement définie au regard de la capacité de *dire* son état mental, elle est, tout comme les théories représentationalistes précédentes, dans l'incapacité d'attribuer une forme de conscience à des êtres qui ne sont pas doués de langage. Une approche comme celle de Campbell permet certes d'étendre les modalités d'expression rationnelle au-delà de la verbalisation (avec le cri, par exemple), mais elle ne rend pas compte du fait plausible qu'un animal ou un jeune enfant puissent éprouver *consciemment* une sensation, percevoir ou croire *consciemment*, sans pouvoir l'exprimer rationnellement au sens où elle l'entend. Or ce point est tout sauf anecdotique, car il semble fondamental pour distinguer les capacités de réponse rationnelle susceptibles de constituer une connaissance de celles qui ne le sont pas. Après tout, une machine à café peut certainement, en *un* sens, répondre de façon rationnelle à l'introduction d'une pièce de monnaie en indiquant son état interne sans fondement épistémique (« En attente de préparation »), et le faire, qui plus est, dans un cadre téléologiquement déterminé par son programme. Pourtant, peu de philosophes seraient prêt à accorder un pouvoir de connaître à une machine à café. Mais il devient difficile de distinguer la capacité de réponse rationnelle de la machine de celle d'un sujet connaissant sans faire appel à des notions épistémiques que la théorie vise à expliquer. Il semble qu'il faille supposer au moins ce que nous avons appelé une conscience intransitive et celle-ci, comme on l'a vu, ne se réduit pas à la possession d'états phénoménaux que l'on pourrait refuser à la machine¹⁰. L'expressiviste répondra que la machine ne peut pas *exprimer* son état interne, mais seulement en faire un compte rendu. Sans doute, mais un comportement expressif n'est pas nécessairement conscient. Pour rendre compte de ce qui nous distingue de la machine d'un point de vue épistémique, il nous faudrait donc un critère non-circulaire d'une expression *consciente*, ce qui n'est pas une mince affaire si l'on définit, comme les expressivistes, la conscience en termes de capacité d'expressivité.

4.2. Le réflexivisme sartrien de Boyle

L'autre approche de la thèse de l'équivalence apporte une solution à certaines de ces difficultés. Selon elle, toute assertion portant sur le monde suppose une compréhension tacite de l'attitude dans laquelle le sujet est engagé à l'égard du monde. Juger « *p* » suppose implicitement l'auto-attribution « Je crois que *p* » et c'est la raison pour laquelle la phrase

¹⁰ Une position comme celle du panpsychisme pourrait néanmoins conduire à lui attribuer une forme de conscience phénoménale (Chalmers 2010).

à la manière de Moore (« Je crois que p , mais non p ») est paradoxale, mais cette fois-ci, sans être une contradiction directe : elle affirme que je crois que p tout en suggérant implicitement que je ne le crois pas dans la seconde partie. Dans cette perspective, l'auto-attribution est une explicitation de cet élément tacite dans un énoncé réflexif. C'est l'approche réflexiviste. L'idée générale de cette approche pourrait permettre d'éclairer le principe de transparence. Boyle (2024) a défendu récemment une thèse de ce type¹¹ : la connaissance de soi transparente s'explique en vertu du fait que nous possédons une connaissance *tacite* de la façon dont nous représentons le monde (sur le mode de la croyance, de l'intention, de la perception, etc.). Tout ce que le sujet a à faire, c'est rendre explicite, par la réflexion, ce qu'il sait déjà de façon implicite. Une telle explicitation ne nécessite pas de lui une inférence, mais seulement un changement de focalisation de son attention, depuis le monde envers lequel il est engagé vers son engagement avec lui.

Il convient alors de distinguer cette connaissance tacite d'une forme d'auto-représentation afin d'éviter les apories évoquées précédemment (2.3). Boyle (2024, chap. 3) propose pour cette raison une version du réflexivisme qui s'appuie sur l'idée sartrienne de conscience non-positionnelle. En vérité, lorsque je pense « Il pleut », le « Je crois que » n'est pas *représenté* implicitement dans un *contenu* secondaire, comme dans la théorie de l'auto-représentation, mais seulement *présupposé* dans le *mode de présentation* du contenu, c'est-à-dire la manière spécifique de le penser¹². Et il en va de même pour tous les cas de transparence. Par exemple, dans un énoncé perceptif comme « Ce chat ronronne », le démonstratif suppose un mode perceptuel de présentation, puisqu'il n'a ce sens perceptif plutôt qu'anaphorique que si le chat est présenté comme disponible à la monstration au moment où l'énoncé est proféré ou pensé. Mais le contenu de l'énoncé ne suppose pas le *concept* de perception, comme si celui-ci en faisait implicitement partie. Il n'est pas davantage à propos de celui qui perçoit. L'énoncé ne signifie pas « Je perçois un chat qui ronronne », mais comprend implicitement un mode perceptuel de présentation du chat. De même pour l'intention, quand je pense « Je vais φ », la signification de cette pensée n'est pas « Je vais φ en vertu de l'intention que j'ai maintenant de φ ». En vérité, en pensant « Je vais φ », je ne m'attribue pas une intention, mais je pense que je ferai φ d'une façon qui présuppose implicitement une telle intention. L'intention n'est pas représentée, mais elle est tout de même consciente. Cette conscience s'exprime notamment dans les types spécifiques de justifications et de conséquences que j'envisage pour la proposition « Je vais φ ». Par exemple, si je la fonde sur des justifications pratiques plutôt que théoriques, si je considère que φ devrait être fait, selon le rapport que j'établis entre la proposition « Je vais φ » et mes autres plans, etc. Bref, la conscience tacite de mon intention ne réside pas dans une auto-attribution explicite, mais dans ce que je fais ou peux faire de la proposition « je vais φ », dans la façon, dit Boyle, dont je suis susceptible de « *traiter* » cette proposition. C'est là que se joue la différence entre une intention et une prédiction théorique concernant mon futur, quand bien même cette prédiction serait-elle reliée à mon intention présente (2024, 69-70).

Dans le cas de la croyance, enfin, l'assertion « Il pleut » ne signifie pas « Je crois qu'il pleut ». Néanmoins, la conscience non-positionnelle de la croyance est présente dans le mode assertorique selon lequel le sujet envisage la proposition, plutôt que, par exemple, sur un

¹¹ Sa position n'est pas justifiée à partir d'une analyse du paradoxe de Moore, mais seulement de la transparence. Néanmoins comme je l'ai déjà dit, transparence et paradoxe de Moore sont liés.

¹² Voir aussi O'Brien 2007, chap. 6 et Recanati 2010.

mode seulement hypothétique (il traite la proposition comme « fermée », dit Boyle, à l'égard de la délibération).

En revanche, l'auto-attribution explicite d'une perception, d'une intention ou d'une croyance est possible pour un sujet qui possède les concepts psychologiques correspondants. La réflexion consiste justement à expliciter la conscience non-positionnelle dans un jugement d'auto-attribution. C'est une explicitation de ce qui est compris, non dans le contenu, mais dans le mode¹³. Si je crois qu'il pleut, le contenu de ma croyance ne comprend aucune référence à moi-même ni à mon attitude. Mais cette attitude est caractérisée par un mode de présentation dans lequel s'exerce une conscience préréflexive de ma croyance. C'est ainsi que la transparence se trouve expliquée : dire ou penser « *p* » comprend implicitement « je crois que *p* » et le passage de l'attitude tournée vers *p* à la conscience positionnelle ou transitive de cette attitude est une explicitation réflexive rendue possible par la maîtrise du concept de croyance. Cette thèse aurait plusieurs avantages : 1) elle permettrait d'attribuer des croyances (ainsi que des intentions, etc.) à des créatures ne disposant pas des concepts psychologiques correspondants ; 2) elle rendrait compte d'une authentique connaissance en première personne, car, par la réflexion, je n'aurais pas à apprendre quelle est mon attitude sur la base de preuves, puisque je suis déjà censé en avoir implicitement conscience ; 3) elle pourrait permettre de rendre compte de l'articulation des usages (i) et (iii) de « Je crois que », en disant que (iii) est implicite dans (i) – bien que Boyle n'en parle pas.

Le problème de cette approche, néanmoins, c'est que la nature de cette conscience tacite, préréflexive, n'est pas claire. Si, comme le dit Boyle, « ce qui justifie notre auto-attribution réflexive est une conscience de l'état mental pertinent » (Boyle 2024, 74)¹⁴, comment comprendre cette justification ? De deux choses l'une.

1) Ou bien cette conscience tacite n'est pas elle-même déjà une connaissance de mon état mental. Elle est alors quelque chose qui conditionne mon rapport au monde et que je peux connaître si j'y prête attention. Elle sert ainsi de base de justification à mon auto-

¹³ Boyle donne à l'expression « mode de présentation » un sens qui n'est ni fregéen ni néo-russellien, c'est-à-dire ni une propriété du contenu, ni seulement une propriété du mode de relation au contenu : « Je considère qu'une caractéristique d'une représentation appartient à son mode de présentation de son objet, plutôt qu'au contenu présenté, lorsque cette caractéristique exprime un aspect de la relation cognitive du sujet à l'objet représenté, mais d'une manière qui ne peut être saisie en commençant par un type générique de contenu représentationnel et en ajoutant d'autres spécifications portant sur la relation cognitive du sujet à cet objet. » (Boyle 2024, 75). Il me semble que l'on peut interpréter cette idée comme signifiant que le mode de présentation ne fait pas partie du contenu, mais affecte la présentation du contenu. Par exemple dans « Ce chat ronronne », le démonstratif perceptuel fait partie de la proposition, mais pas l'idée de perception elle-même. En pensant « *I will φ* » l'auxiliaire « *will* » peut servir soit d'auxiliaire d'intention (noté « *I will i* » par Boyle), soit d'auxiliaire de prédiction (noté « *I will bf* », pour *belief*). Lorsqu'il s'agit d'une intention, c'est le *will i* qui est pensé dans le contenu. Pour la croyance, c'est semble-t-il, le mode assertorique qui affecte le contenu. Les modes de présentation sont ainsi les manières caractéristiques par lesquels différents types d'états mentaux représentationnels présentent leur contenu (p. 97) : « les modes de présentation sont des choses liminaires, qui chevauchent la frontière entre le connaisseur et l'objet connu. D'une part, en tant que modes de présentation d'un *objet*, ils affectent la manière dont cet objet est présenté. D'autre part, en tant que modes de *présentation* d'un objet, ils impliquent les capacités cognitives du sujet » (p. 217).

¹⁴ Boyle conçoit l'auto-attribution comme l'expression d'une réflexion du sujet sur son état mental. Sur ce point, il n'est pas certain qu'il soit fidèle au Sartre dont il se revendique. Selon ce dernier, en effet, les réponses spontanées à la question « Que faites-vous ? », par exemple, « ne nous transportent pas sur le plan de la réflexion » (Sartre 2003, 71-72). De ce point de vue, Sartre est plus proche de l'expressivisme.

attribution réflexive. Mais en quoi ma réflexion diffère-t-elle alors d'une sorte d'auto-observation spectatorielle ? Je *constate* simplement que je possède une conscience implicite et je l'explicite. Par analogie, si je porte des lunettes, celles-ci sont bien une condition transparente de ma vision des objets. Mais si je veux en connaître quoi que ce soit, il n'y a rien à faire, il faut que je les regarde comme les autres objets, que j'opère des inférences à partir d'indices accessibles dans mon champ visuel ou que je m'informe auprès d'autrui. Selon cette option, la question est encore celle des bases de justification épistémique. De ce fait, la position du sujet à l'égard de son état mental est spectatorielle. Il en va comme si on lui demandait « Comment savez-vous que vous croyez que *p* ? » et qu'il répondait « Parce que j'en ai une conscience implicite. » ; ce qui revient à prendre la question au sérieux (Roessler 2024). Mais une telle question ne peut être prise au sérieux que si l'on suppose une forme d'aliénation.

2) Ou bien la conscience non-positionnelle est déjà une forme de connaissance de son état mental. Mais alors comment cette conscience non-positionnelle peut-elle être tacite si ce n'est seulement au sens, sans conséquence, qu'elle n'est pas encore explicitement verbalisée ? Si elle est véritablement tacite, comment peut-on dire que le sujet sait déjà (avant la réflexion) ? En effet, comme Boyle le reconnaît lui-même, une connaissance est individuée par son contenu propositionnel (ma croyance que *p* est individuée par *p*), mais si celui-ci est tacite comment peut-il l'individer ? Alors de quoi est-elle connaissance ? Il faudrait proposer une autre forme, non-représentationnelle, de connaissance pour en rendre compte. Mais en concevant la conscience intransitive comme une conscience implicite *pré-réflexive*, Boyle la pense toujours à l'aune de la réflexion et donc de la représentation.

5. Conscience intransitive et savoir-faire

5.1. Conscience intransitive et capacité de réactivité intelligente

Il convient donc de proposer un modèle non-représentationnel de la conscience intransitive qui en fait néanmoins une forme de savoir. Mais quel modèle ?

Il y a peut-être, dans l'analyse de Boyle lui-même, une indication pour un tel modèle lorsqu'il évoque la connaissance en première personne de nos intentions. Il semble en effet suggérer que la connaissance que l'agent possède de son intention réside, non dans le contenu de cette intention, mais dans ce qu'il *fait* et est *capable* de faire de ce contenu. Ici, la conception que propose Boyle de la conscience *pré-réflexive* rencontre l'approche capacitaire de l'expressivisme. De fait, Boyle en vient à caractériser la conscience implicite qu'un agent possède de son état mental *M* en termes de « réactivité intelligente à ce qui est en fait notre être en *M* » (2024, 98). Cette idée rappelle la thèse de Campbell, selon laquelle la connaissance de soi est une capacité de réponse rationnelle expressive au fait d'être dans *M*, mais repensée en termes de *mode de présentation*, donc de *manière de penser* – et donc, pourrait-on ajouter, de façon *adverbiale*.

Néanmoins, cette remarque de Boyle, tout comme l'analyse de Campbell, font référence à des capacités de *second ordre*, c'est-à-dire concernant *le fait d'être* dans un état mental donné. Mais si on prend au sérieux l'idée de transparence, comme le fait Boyle, alors la conscience intransitive devrait consister directement dans une capacité de réagir intelligemment au *contenu* de cet état mental. C'est une conséquence directe de la critique de la thèse des représentations d'ordre supérieur. Si la conscience intransitive ne résulte pas de la représentation d'un état mental dans un autre, elle doit être une propriété *inhérente* aux états mentaux de premier ordre, c'est-à-dire inhérente au fait pour le sujet

d'être dans un tel état mental. Pourquoi faudrait-il alors supposer en plus comme nécessaire, une capacité de second ordre par rapport à son état mental ?

Par ailleurs, croire, percevoir ou avoir une intention, ce n'est pas représenter passivement un état de chose, un objet ou une action. Ces attitudes sont chacune constituées d'un ensemble de tendances (capacités et dispositions) interconnectées et d'exercices de ces tendances, à la fois actives et réactives. Adopter le point de vue de quelqu'un qui croit que p , a l'intention de φ ou perçoit O , c'est exercer et être capable d'exercer certaines tendances relativement à un contenu.

Il appert alors que si la conscience intransitive est bien une propriété inhérente à une attitude mentale, les capacités constitutives de cette attitude, ou du moins certaines d'entre elles, doivent être aussi les capacités constitutives de la conscience intransitive propre à cette attitude. Il n'y a pas à supposer une relation supplémentaire du sujet à ces capacités pour rendre compte du caractère conscient de l'attitude qu'elles définissent.

Une telle idée se trouve peut-être déjà chez Aristote lorsqu'il dit que « l'esprit se comprend par participation à ce qu'il comprend » (1072b 19-20) – sans doute l'une des premières formulations du principe de transparence. Comme le notent Anscombe et Geach dans leur explication de cette citation, quand nous voulons savoir si quelqu'un a compris un théorème, nous nous en assurons en observant si, dans ce qu'il dit quand il explique le théorème, nous trouvons bien le théorème, « car c'est le théorème que nous lui demandons d'exposer, et non quelque chose d'autre qui pourrait s'appeler la compréhension du théorème », « et qui serait dans une certaine relation avec le théorème ». Expliquer sa compréhension du théorème, ce n'est rien d'autre qu'expliquer le théorème, puisque l'avoir compris, c'est être capable de l'expliquer. A titre de corollaire, ajoutent Anscombe et Geach, celui qui comprend le théorème « comprend aussi qu'il le comprend », puisque si nous sommes assurés qu'il comprend le théorème, nous n'aurons pas à lui demander en plus « Comprenez-vous *que* vous comprenez ? Pouvez-vous nous expliquer, non pas ce que vous comprenez, mais la compréhension que vous en avez ? ». Si l'élève a compris, l'explication qu'il donne du théorème « est accessoirement une explication de la compréhension qu'il en a » (Anscombe & Geach 2014, 57).

De façon analogue, si nous nous assurons qu'une personne possède une certaine croyance, parce qu'elle nous le dit ou se comporte d'une façon appropriée, alors nous n'aurons pas à lui demander en sus de nous expliquer la compréhension qu'elle possède du fait qu'elle croit. L'expression de ce qu'elle croit est accessoirement une expression de la compréhension qu'elle a du fait de le croire. De même que la capacité à expliquer le théorème est constitutive de la compréhension que l'élève possède de sa propre compréhension, de même l'ensemble des capacités constitutives de sa croyance que p est constitutif de sa conscience intransitive. Autrement dit, ces capacités ne sont pas ici des capacités de second ordre portant *sur* le fait de posséder une croyance (ce qui nous renverrait encore à une forme de conscience transitive). Croire consciemment que p , c'est exercer et être capable d'exercer certaines capacités constitutives de la croyance.

L'exemple de la compréhension est intéressant également à un autre égard. Le verbe « comprendre » admet deux sens, l'un factif, l'autre non. Si l'explication de l'élève bénéficie bien de l'autorité de la première personne pour ce qui concerne la façon dont il a compris, au sens non-factif, le théorème, elle ne fait pas autorité concernant le fait d'avoir compris, au sens factif, le théorème. Mais le principe de transparence s'applique toujours, même sans l'autorité, si l'explication que l'élève donne de sa compréhension du théorème n'est autre que l'explication qu'il donne du théorème. Il y a, par contraste, une façon dont l'élève pourrait manifester une relation aliénée à sa compréhension. Il pourrait répondre à la demande de justifier sa compréhension en déclarant « Mon professeur m'a dit que j'avais

compris ». Au contraire, rendre compte, *en première personne*, du fait d'avoir compris, c'est exposer *ce qui* est compris et non pas parler *de* sa compréhension¹⁵. Il y a donc bien des cas où la transparence vaut sans l'autorité. On voit ainsi qu'il n'y a pas de corrélation systématique entre le fait qu'un état mental soit authentiquement en première personne, c'est-à-dire qu'il ne soit pas aliéné, et le fait qu'il y ait autorité de la première personne. L'élève comprend consciemment s'il est capable de rendre compte *en personne* des capacités constitutives de la compréhension du théorème, de donner sens par lui-même à la proposition comprenant le théorème. De la même façon, croire consciemment, c'est être capable d'exercer *en personne* certaines des capacités et dispositions constitutives de sa croyance. Il faut alors préciser ce que signifie ici « en personne ». Comme on va le voir, l'expression ne signifie pas seulement « à l'exclusion d'un tiers », comme dans l'exemple du professeur, mais également « de façon holistique » par distinction avec « subpersonnelle ».

5.2. Conscience et intégration fonctionnelle

Laissons d'abord de côté la transparence et de revenons à l'exemple de la douleur. Un certain nombre de philosophes, à commencer par Wittgenstein et Sartre, considèrent qu'il n'y a pas de différence entre avoir une douleur et éprouver consciemment une douleur. Il n'y aurait pas de sens à pouvoir ignorer sa douleur, car une douleur que l'on ignore est une douleur que l'on ne ressent pas et n'est donc pas une douleur. D'autres, comme les partisans des représentations d'ordre supérieur, considèrent au contraire qu'il existe des douleurs inconscientes. Je pense qu'il y a une façon de réconcilier ces deux opinions. Une douleur consciente est une douleur que l'on peut attribuer à l'organisme lui-même comme un tout au regard d'un ensemble de fonctions intégrées¹⁶ et coordonnées constitutives de l'organisme qu'il est. Supposons au contraire qu'une certaine information soit traitée seulement au niveau subpersonnel du système nerveux de l'organisme, sans connexion possible avec ses capacités à réagir à la douleur, non seulement de façon externe, par un comportement de fuite, mais également de façon interne, par la tendance ou le désir de ne plus l'éprouver. Il me semble qu'on pourrait appeler cela une douleur lichtenbergienne, c'est-à-dire sans sujet¹⁷. On pourrait dire « Il y a de la douleur » mais sans pouvoir dire que *l'organisme* lui-même la ressent, car il n'y aurait tout simplement pas de sens à l'attribuer à l'organisme compris comme un système de fonctions intégrées. On pourrait peut-être bien parler de douleur inconsciente, si l'on entend par là qu'une sensation est traitée seulement à un niveau subpersonnel, mais ce ne serait pas alors une douleur inconsciente *de l'organisme* lui-même. Dans ces conditions, on voit également que l'aspect qualitatif ou phénoménal d'une information ne suffit pas à rendre compte de son caractère conscient, il faut encore qu'elle soit intégrée aux fonctions et capacités globales du sujet – à moins de considérer que la conscience puisse être une propriété d'un état mental indépendante d'un sujet, ce qui est douteux¹⁸. Notons, néanmoins, qu'une façon de réagir

¹⁵ Ce point fait écho au problème de la factivité évoqué précédemment en rapport avec l'expressivisme grammatical et lui donne raison en partie. Le fait que l'usage (i) de « Je crois que » puisse être quasi-factif n'exclut pas qu'il soit proprement en première personne. Reste néanmoins le problème de l'articulation avec (iii).

¹⁶ Sur la notion d'intégration, voir Hyman 2015, chap. 2.

¹⁷ Un fonctionnalisme strict, me semble-t-il, devrait nier ici l'existence même d'une douleur. Cela ne me semble pas évident.

¹⁸ Pour les raisons ici exposées, je ne pense pas que l'on puisse identifier phénoménalité et conscience (comme le veulent les théories phénoménologiques), ni réduire la phénoménalité à une capacité ou un savoir-faire (comme le veut Lewis 1988).

intelligemment à la douleur pourrait être aussi d'essayer de ne pas y prêter attention, comme un sportif blessé pendant un match. Dans ce cas, en supposant que le sportif y parvienne, il y a toujours un sens à lui attribuer une douleur même quand il ne la ressent plus. Peut-être est-ce un cas où l'on serait alors prêt à parler de douleur inconsciente *du sujet lui-même*¹⁹.

De façon analogue, pour un état mental marqué par la transparence, comme la croyance, il ne suffit pas qu'une certaine information soit traitée par le système cognitif du sujet pour en faire une croyance du sujet. Encore faut-il qu'elle puisse être traitée par celui-ci au regard de fonctions intégrées de haut niveau qui définissent les capacités de réaction intelligente de ce sujet comme un tout unifié. Pour ce faire, il faut regarder pour ainsi dire *au-delà* de la simple présence de l'information dans le système cognitif pour considérer le contexte plus large qui correspond à ce que le sujet fait et peut faire de cette information. On lui attribuera une croyance dans la mesure où on l'estime capable de réagir à l'information en lui donnant sens en pratique et cette capacité est également ce qui définit son caractère conscient. Cette façon de donner sens à une information dépend d'un ensemble de contraintes fonctionnelles et rationnelles qui sont propres à un individu de telle ou telle espèce.

Si l'on estime alors qu'il existe des croyances inconscientes attribuables à l'individu, c'est parce qu'on considère qu'elles satisfont un certain nombre de contraintes fonctionnelles et rationnelles de niveau holistique, mais que certaines d'entre elles manquent alors même que nous les tenons pour paradigmatiques ou essentielles. Par comparaison, on peut dire qu'il est essentiel à un canard d'avoir seulement deux pattes, même si certains canards en ont trois²⁰. De façon analogue, on dira que le paradigme de la croyance est une croyance consciente, mais il se peut qu'il existe aussi des croyances inconscientes. Cette possibilité est cependant insuffisante pour que la conscience intransitive ne soit pas une propriété *générique* des croyances, c'est-à-dire une propriété définitionnelle qui n'est pas pour autant une propriété nécessairement instanciée dans *tous* les cas. Dans le cas définitionnel – par rapport auquel chaque croyance est une croyance – la croyance est consciente. C'est à l'aune de ce cas définitionnel, avec lequel elles partagent néanmoins certaines propriétés paradigmatiques, que certaines dispositions inconscientes pourraient éventuellement être considérées comme des croyances même si elles ne lesinstancient pas toutes. En conséquence, c'est le caractère inconscient de la croyance qui demandera une explication *par soustraction* de capacités, plutôt que son caractère conscient qui demandera une explication par un mécanisme *additionnel* de détection interne, comme le veulent au contraire les théories représentationalistes. Une croyance inconsciente – si l'on admet cette notion – est une croyance avec quelque chose en moins, plutôt qu'une croyance consciente, une croyance inconsciente avec quelque chose en plus.

5.3. La conscience intransitive comme savoir-faire

On peut donc supposer qu'il existe une forme de conscience intransitive qui est inhérente au mode de présentation caractéristique d'une attitude mentale (Boyle) et que ce mode de présentation se comprend relativement à un ensemble de capacités fonctionnelles et rationnelles interconnectées. Il reste alors à expliquer à la fois ce qui en fait une forme de *savoir* et dans quelle mesure celui-ci n'est pas *représentationnel*.

¹⁹ C'est la différence entre « inconscient » et « subpersonnel ».

²⁰ Voir, par exemple, « Gers : un canard à trois pattes amuse la ferme d'Empluhaut », *La Dépêche* du 26/05/2021.

A elles seules, des dispositions purement fonctionnelles ne suffisent pas à rendre compte d'un savoir. Je soutiendrai donc que ce savoir est une forme de *savoir-faire*. Comme le suggèrent certaines remarques de Boyle, c'est dans la façon dont un sujet *traite* un contenu de pensée que réside la conscience intransitive. Habiter le point de vue conscient de quelqu'un qui croit que p , c'est être familier d'un ensemble d'actions et de réactions constitutives d'une attitude doxastique à l'égard de p . Le sujet est alors comme familiarisé avec un ensemble de contraintes de rationalité constitutives d'un état mental de type croyance qu'il sait appliquer sans avoir nécessairement à les représenter autrement que dans son comportement. Il sait y faire avec un contenu qu'il présente sur un mode doxastique. L'une de ces contraintes de rationalité est l'interdit d'attitudes conjointes à l'égard d'un contenu qui pourraient s'énoncer dans une phrase paradoxale comme « Je crois que p , mais non- p » (contrainte de consistance). Mais cette contrainte peut manquer lorsqu'une croyance est aliénée. C'est notamment dans ce type de cas que l'on sera tenté de parler de croyance inconsciente.

On comprend par là également que ce serait une erreur de chercher une modalité unique de la conscience intransitive qui s'exercerait de façon identique pour chaque type d'état mental. Les contraintes rationnelles liées à la croyance ne sont pas les mêmes que celles de l'intention, par exemple, ou du désir²¹. En vérité, la conscience intransitive propre à une attitude se définit de façon *différentielle* par rapport aux autres types d'attitude possibles. En se comportant d'une certaine façon à l'égard de p , le sujet dessine les contours d'une attitude doxastique et la conscience intransitive de cette attitude réside dans un ensemble de comportements possibles qui sont discriminants à l'égard d'autres attitudes. C'est en exerçant les capacités et les contraintes rationnelles constitutives de la croyance, par distinction, par exemple, avec celles de l'intention ou du désir – mais aussi de la conjecture, de l'interrogation ou de la supposition, etc. – que le sujet adopte le point de vue conscient de quelqu'un qui croit que p . Mais ce serait faire preuve d'un intellectualisme excessif que de supposer que j'aie besoin de me représenter le fait que je crois que p pour exercer consciemment les contraintes de rationalité appropriées.

A la différence de la croyance, le désir, par exemple, n'est pas soumis à l'interdit d'un comportement exprimable dans un énoncé comme « Je désire que p , mais non- p », pas plus qu'à l'exigence de cohérence entre nos désirs. En revanche, il l'est sans doute par rapport à un énoncé comme « Je désire que p , mais p n'a rien de bon ». L'intention admet un analogue du paradoxe de Moore (« J'ai l'intention de φ , mais je ne le ferai pas »), mais à la différence d'une croyance, elle peut perdurer rationnellement au-delà du constat de son insatisfaction (contrainte de stabilité). Je peux constater que j'ai échoué à nouer mes lacets sans abandonner mon intention de les nouer. Un tel abandon serait même généralement considéré comme irrationnel. Mais si je crois que je les ai noués et que je constate que ce n'est pas le cas, c'est une raison d'abandonner ma croyance²².

Un des intérêts de cette hypothèse sur la conscience intransitive, c'est qu'elle permet aussi de relativiser les modalités et les degrés de cette conscience en fonction des capacités des espèces auxquelles nous sommes prêts à attribuer des états conscients. Cette conscience dépend, pour chaque espèce, des fonctions intégrées au niveau global constitutives de l'individu. Certaines de ces fonctions étant acquises ou développées au cours de

²¹ Voir Bratman 1987.

²² Cette thèse permet également d'imaginer des cas où l'agent hésite dans l'application des contraintes de rationalité appropriées : s'agit-il d'une croyance ou d'un désir ? Il ne sait pas vraiment comment s'y prendre à l'égard d'un contenu. *Parfois*, il aura alors à interpréter son attitude ou à en décider. Pour ce faire, il se référera à un paradigme, raison pour laquelle ce cas ne peut pas être la règle.

l'apprentissage, on peut les rattacher à l'idée de savoir-faire. On considérera alors que, même si les croyances, les perceptions, les désirs, les intentions ou les sensations occupent des places dans les formes de vie des chiens ou des chats qui sont semblables à bien des égards à celles qu'elles occupent dans la nôtre, ces états mentaux ont néanmoins chez nous une dimension conceptuelle qui définit de façon spécifique la conscience intransitive qui leur est propre. Ainsi, la capacité d'exprimer une croyance, non seulement dans un énoncé du type « Il pleut », mais dans une forme auto-attributive comme « Je crois qu'il pleut » est certainement une des capacités constitutives du mode propre à une croyance pour un être humain ayant atteint un certain degré de maturité linguistique. Croire que *p*, c'est une manière de traiter *p* selon des contraintes rationnelles et fonctionnelles qui sont immanentes à un comportement relatif à *p*, et qu'un sujet doué des capacités conceptuelles requises peut reconnaître, parce qu'il a appris à les appliquer en même temps qu'il a appris à les conceptualiser sous formes d'expression en première personne et de descriptions en troisième personne. En disant « J'ai mal » ou « Je crois que *p* » je ne fais pas qu'*exprimer* mon état mental, je le *communique* dans une auto-attribution par laquelle je me situe d'emblée dans un monde commun et partagé, au sein duquel ce que je ressens en souffrant ou ce que je crois est aussi ce qu'autrui peut ressentir ou croire. En suivant une suggestion de Johannes Roessler (2024, note 13), dans sa recension critique de Boyle (2024), on peut dire que s'il y a bien une connaissance implicite, comme le dit Boyle, c'est peut-être alors d'une façon analogue à la connaissance grammaticale de quelqu'un qui sait en pratique utiliser un certain nombre de règles syntaxiques sans avoir appris à les théoriser. Il sait en pratique distinguer un accusatif d'un nominatif sans pour autant pouvoir objectiver la règle, parce qu'il ne dispose pas des concepts correspondants. Mais, comme le note Dummett (1993, chap. 3), un tel sujet linguistique est capable de reconnaître immédiatement la règle qu'il utilisait lorsqu'on la lui enseigne. Il y a donc bien ici aussi une forme d'intrication entre un savoir-faire et un savoir conceptuellement informé.

On pourrait ainsi rendre compte, au regard de l'apprentissage du langage, de l'articulation entre un usage de « Je crois que *p* » où l'expression est pragmatiquement et sémantiquement équivalente à « *p* » (celui sur lequel se focalise l'expressiviste grammairien) à un usage auto-attributif non-aliéné²³. L'enfant apprend d'abord, dans une phase égocentrée, à utiliser l'expression « Je crois que » comme équivalent de l'assertion du contenu de sa croyance et acquiert peu à peu une capacité d'auto-attribution par laquelle il se situe en rapport avec le point de vue d'un interlocuteur possible. On comprend en ce sens qu'il y a un lien étroit, comme le soutient le réflexivisme de Boyle, entre la conscience intransitive et la conscience transitive, réflexive, de nos états mentaux conscients, sans qu'il faille renoncer à distinguer un usage proprement réfléchi de l'auto-attribution et un usage expressif qui, comme dirait Sartre, ne nous transporte pas « sur le plan de la réflexion²⁴ ».

5.4. Objections et réponses

On pourrait néanmoins vouloir objecter que les attitudes mentales visées dans cet article ne satisfont pas certaines propriétés du savoir-faire. J'énonce une à une les objections et les réponses.

Croire, avoir une intention, etc. ne sont pas des actions.

²³ Respectivement, usage (i) et (iii) ci-dessus (4.1).

²⁴ Voir supra, note 13.

Il ne s'agit pas de dire que croire est une *action*, une seule action et *seulement* une action. « Croire » est un terme générique qui réfère à un ensemble de *tendances*, de l'ordre de dispositions et de capacités, unies par des contraintes rationnelles, et il en va de même pour les autres attitudes mentionnées. Ces tendances s'expriment notamment dans des actions (comme prendre mon parapluie, lorsque je crois qu'il pleut ou dire « Il pleut »). Il y a cependant aussi une part des capacités de réactivité intelligente à un contenu, qui sont des capacités émotionnelles. Par exemple, ma capacité à me réjouir d'un contenu de croyance n'est pas à proprement parler une capacité à agir. La conscience intransitive réside dans le fait de savoir exercer ces tendances selon les contraintes de rationalité discriminantes qui sont constitutives d'une attitude.

On conçoit souvent la croyance ou l'intention comme des dispositions plutôt que des capacités (Ryle 2005). S'il en est ainsi, le point de vue authentiquement en première personne devient problématique, car le sujet devra observer son comportement ou faire des inférences contrefactuelles pour s'attribuer de telles dispositions (Wittgenstein 2004, 272).

Il convient en effet de distinguer disposition et capacité. Je ne nie pas que parmi l'ensemble des tendances constitutives d'une attitude mentale il puisse également y avoir des dispositions. C'est peut-être ce que suggère la possibilité de dire, par exemple, qu'une intention est résolue ou qu'une croyance est ferme, alors qu'une capacité, comme celle de jouer du piano, n'admet pas de telles qualifications. Mais il se peut qu'en qualifiant ainsi ses attitudes, on entende signifier que l'agent est résolument ou fermement disposé à exercer les capacités constitutives de l'intention ou de la croyance. C'est précisément dans cette distinction entre capacités et dispositions que se joue la possibilité d'une auto-attribution authentiquement en première personne. Si je sais jouer du piano, je n'ai pas à m'observer pour savoir que je joue du piano quand j'en joue. J'en ai une connaissance pratique, c'est-à-dire justement ce type de connaissance qui *consiste* dans l'exercice d'un savoir-faire (Anscombe 2002, § 48).

L'exercice d'un savoir-faire est intentionnel, mais croire, avoir une intention etc. ne sont pas des comportements intentionnels. Dans le cas contraire, il y aurait nécessairement une régression.

J'admets qu'agir intentionnellement suppose la maîtrise d'un savoir-faire. Mais je conteste la réciproque. En écoutant quelqu'un parler, j'exerce un grand nombre de compétences rationnelles acquises grâce auxquelles je peux comprendre ce qui est dit. Je ne les exerce pas pour autant intentionnellement, si ce n'est au sens où je pourrais intentionnellement ne pas écouter. Mais dans le même sens, je pourrais m'empêcher de croire, de former une intention ou de percevoir. L'exercice d'un savoir-faire relève d'un comportement *intelligent*, mais pas nécessairement intentionnel (Ryle 2005 ; Hornsby 2011). On peut d'ailleurs croire, vouloir, désirer, mais aussi percevoir, de façon plus ou moins intelligente, tout comme jouer du piano.

L'exercice d'un savoir-faire suppose une possibilité d'erreur quant au fait que l'on exerce effectivement ce savoir-faire (Anscombe 2002). Mais l'autorité en première personne semble invalider cette possibilité pour un certain nombre d'états mentaux conscients.

J'ai déjà montré qu'un jugement authentiquement en première personne n'implique pas *toujours* l'autorité²⁵, mais certains verbes, comme « croire » ou « vouloir », etc. semblent néanmoins supposer par défaut une telle autorité. Cependant, il existe aussi des formes de savoir-faire qui sont soustraites dans le même sens à la possibilité de l'erreur. Attendre ou chercher, sont des actions intentionnelles qui supposent des compétences rationnelles acquises. Mais je ne peux pas me tromper sur le *fait* que j'attends ou que je cherche. Je peux certes me tromper en n'attendant pas au bon endroit ou la bonne personne, en ne cherchant pas là où il faut ou correctement, mais tout comme je peux me tromper en croyant ou formant une intention.

On sait qu'une interprétation intellectualiste du savoir-faire gagne de plus en plus en importance (Stanley & Williamson 2001 ; Stanley 2011 ; Pavese 2015 ; 2017). Selon cette approche, savoir-faire ϕ est une façon de connaître une proposition comprenant des manières appropriées de réaliser ϕ . Le savoir-faire serait donc de type représentationnel. En conséquence, les règles de rationalité et les règles fonctionnelles qui définissent à la fois une attitude mentale et la conscience intransitive qui lui est inhérente pourraient être comprises comme des prescriptions comportementales que le système cognitif et moteur de l'individu devraient représenter. La conscience intransitive ne serait donc pas réellement intransitive, mais bien représentationnelle.

Il me semble néanmoins que l'argument de la régression complété par le principe de consciencité (*supra*, 2.2) permet de répondre à cette objection et constitue par la même occasion un contre-argument à l'idée que toute forme de savoir-faire est de type représentationnel. En effet, si l'on adopte une version intellectualiste de l'approche capacitaire de la conscience intransitive, le caractère conscient d'un état mental devra tenir du fait qu'il est *représenté* dans un ensemble de prescriptions pratiques constitutives de cet état. On en revient donc à une forme de théorie des représentations d'ordre supérieur. Pour être conscient de son état mental, le sujet devra se représenter cet état sous forme de règles procédurales constitutives. Mais, pour être conscient de quelque chose, il faut être dans une attitude consciente (principe de consciencité). Il faudra donc que cette représentation procédurale soit elle-même consciente en vertu d'une représentation antérieure, etc.

On pourrait vouloir résoudre le problème de la régression en postulant des degrés dégressifs de conscience, la série des représentations procédurales étant comme une chaîne de représentations de moins en moins conscientes. On pourrait également situer la différence entre une conscience spectatorielle de son état mental et une conscience en première personne au regard de cet aspect pratique-prescriptif des représentations procédurales. C'est sans doute une alternative intéressante à l'approche anti-représentationaliste. Mais si le savoir-faire constitutif des attitudes mentales en question ressemble davantage à celui qui est attaché à des actions comme attendre ou chercher, il convient de noter qu'il n'est pas évident que les actions de ce type soient décomposables en un nombre fini d'étapes articulées selon des règles de composition, comme le voudrait l'intellectualisme. Attendre, par exemple, n'est pas réductible à faire les cents pas, lire le journal, et l'ensemble des actions composant une attente particulière²⁶. On peut accomplir toutes ces actions dans le même ordre sans attendre et attendre sans accomplir aucune de ces actions. Et il n'y a pas davantage de raison de penser que les attitudes comme la croyance ou l'intention sont, quant à elles, décomposables. C'est sans doute là ce qui nous

²⁵ Voir *supra*, (5.2) sur l'usage factif de « comprendre ».

²⁶ Je dois cette idée d'action indécomposable à Bruno Gnassounou.

distingue encore de la machine qui, pour sa part, exécute un programme explicable selon un schéma intellectualiste (Pavese 2015). Si nous n'attribuons généralement pas de conscience à la machine, c'est peut-être aussi parce que nous ne lui reconnaissons justement pas de compétences non procédurales.

Conclusion

J'ai tâché de montrer l'irréductibilité d'une forme de conscience intransitive non-représentationnelle. Cette conscience se manifeste, comme le reconnaît Boyle, dans le principe de transparence. Il y a une forme de conscience relative à un état mental intentionnel qui réside dans le mode de présentation de ce qu'il vise, mais qui ne suppose pas elle-même une représentation de cet état mental. J'ai suggéré que cette conscience pouvait être définie en partie en termes de savoir-faire. Croire consciemment que p , c'est traiter p selon un ensemble de tendances constitutives de la croyance et discriminantes à l'égard d'autres types d'attitudes possibles. La conscience intransitive constitutive d'une croyance – ou, ce qui revient au même, le fait de croire consciemment – consiste dans la capacité d'exercer dans son comportement à l'égard d'un contenu, un ensemble de contraintes de rationalité discriminantes par rapport à d'autres attitudes possibles relatives au même contenu. Il est apparu que, si une telle analyse est pertinente, elle est difficilement compatible avec une approche intellectualiste du savoir-faire. Autrement dit, elle exclut le postulat d'une représentation de règles procédurales de rationalité pour rendre compte de ce savoir.

Bibliographie

- Anscombe G. E. M., *L'intention*, trad. M. Maurice et C. Michon, Paris, Gallimard, 2002.
- Anscombe G. E. M. & Geach, P. T., *Trois Philosophes, Aristote, Thomas, Frege*, trad. D. Berlioz et F. Loth, Montreuil-sous-Bois, Ithaque, 2014.
- Bar-On D., *Speaking My Mind: Expression and Self-Knowledge*, Oxford, Oxford University Press, 2004.
- Boyle M., *Transparency and Reflection*, Oxford, Oxford University Press, 2024.
- Bratman M., *Intention, Plans, and Practical Reason*, Cambridge, Harvard University Press, 1987.
- Brentano F., *Psychologie du point de vue empirique*, trad. Gandillac M. Paris, Aubier, 1944.
- Byrne A., *Transparency and Self-Knowledge*, New York, Oxford University Press, 2018.
- Campbell L., « Self-Knowledge : Expression without Expressivism », in *Philosophy and Phenomenological Research*, 104 (1), 2020, 186-208.
- Carruthers P., *Language, Thought and Consciousness*, Cambridge, Cambridge University Press, 1996.
- Carruthers P. « Natural theories of consciousness », in *European Journal of Philosophy*, 6, 1998, 203–2.
- Chalmers D., *L'Esprit conscient. A la recherche d'une théorie fondamentale*, trad. S. Dunand, Paris, Ithaque, 2010.
- Descombes V., *Le Parler de soi*, Paris, Gallimard, 2014.
- Descombes V. & Larmore, C., *Dernières nouvelles du Moi*, Paris, PUF, 2009.
- Dummett M., *The seas of Language*, Oxford, Clarendon Press, 1993.
- Evans G., *The Varieties of Reference*, Oxford, Oxford University Press, 1982.
- Finkelstein D., *Expression and the Inner*, Cambridge, Mass., Harvard University Press, 2003.

Hacker P. M. S., *Human Nature : The Categorical Framework*, Oxford, Wiley-Blackwell, 2007.

Henry M., *L'Essence de la Manifestation*, Paris, PUF 1963.

Horsnby J. « Ryle's Knowing-How and Knowing How to Act », in J. Bengton & M. Moffett, *Knowing How. Essays on Knowledge, Mind and Action*, Oxford, Oxford University Press, 2011, 80–98.

Hyman J., *Action, Knowledge and Will*, Oxford, Oxford University Press, 2015.

Kriegel U., *Subjective Consciousness: A Self-Representational Theory*, Oxford, Oxford University Press, 2009.

Lewis D., « What Experience Teaches », in *Proceedings of the Russellian Society*, 13, 1988, 29–57.

Moran R., *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton University Press, 2001.

O'Brien L., *Self-Knowing Agents*, Oxford, Oxford University Press, 2007.

Peacocke C., « Conscious Attitudes, Attention, and Self-Knowledge », in C. Wright, B. Smith & C. MacDonald (dir.), *Knowing Our Own Minds*, Oxford, Oxford University Press, 1998.

Pavese C., « Practical Senses », in *Philosophers' Imprint*, 15 (29), 2015, 1–25.

Pavese C., « A Theory of Practical Meaning », in *Philosophical Topics* 45 (2), 2017, 65–96.

Recanati F., « Le soi implicite », in *Revue de métaphysique et de morale*, 68, 2010, 475–494.

Roessler J., « Understanding baseless self-knowledge : Boyle on reflection and rationality », in *Philosophisches Jahrbuch*, 131 (2), 2024, 137–154.

Rosenthal D., « Two Concepts of Consciousness », in *Philosophical Studies*, 49 (3), 1986, 329–359.

Rowlands M., « Consciousness and higher-order thoughts », in *Mind and Language* 16 (3), 2001, 290–310.

Ryle G., *La notion d'esprit*, trad. Stern-Gillet S. Paris, Payot, 2005.

Sartre J.P., *La transcendance de l'ego*, Paris, Vrin, 2003.

Sartre J. P., *L'être et le Néant*, Gallimard, 1943.

Stanley J., *Know How*. Oxford, Oxford University Press, 2011.

Stanley J. & Williamson, T., Knowing How, in *Journal of Philosophy* 98, 2001, 411–444.

Wittgenstein L., *Recherches philosophiques*, trad. F. Dastur et alii, Paris, Gallimard, 2004.

Zahavi D., *Self-Awareness and Alterity: A Phenomenological Investigation*, Evanston, Northwestern University Press, 1999.

Remerciements

Pour leur retours précieux et stimulants, m'ayant permis d'améliorer les versions antérieures de ce texte, je tiens à remercier Bruno Gnassounou, Cyrille Michon, Michael Murez, Carole Milin, ainsi que les étudiants de M2 du séminaire de métaphysique de Cyrille Michon à l'Université de Nantes.