

Michael Murez – Putnam et McDowell sur les objets de l'introspection

Le but principal de cet article est de défendre Putnam contre McDowell – ou du moins de défendre certaines thèses et certains arguments du premier dans *Raison, vérité et histoire*¹ (RVH) contre certaines objections du second dans son influent article « Putnam on Mind and Meaning »² (PMM). Puisque par la suite Putnam s'est pour l'essentiel rangé aux positions de McDowell en adoptant le « réalisme naturel »³, il s'agit aussi de défendre Putnam contre lui-même. Cependant, je me focaliserai sur l'échange entre Putnam et McDowell dans les textes susmentionnés, et plus spécifiquement sur un débat qui peut au premier abord sembler étroit, mais qui soulève en fait des enjeux importants concernant la nature de la pensée : le débat sur les objets de l'introspection.

Je commence par présenter le débat sur les objets de l'introspection (1). J'expose ensuite son principal enjeu, du point de vue de McDowell : le lien qui unirait la conception putnamienne des objets de l'introspection comme des véhicules mentaux et sa théorie représentationnelle de la pensée (2). Je clarifie la conception putnamienne des objets de l'introspection, en montrant qu'elle est en fait indépendante de cette dernière (3). J'expose les principaux arguments de Putnam en faveur de cette conception (4), puis les principales objections de McDowell à son encontre (5). Enfin, je réponds à ces objections (6), avant de conclure (7).

1. Objets de l'introspection : véhicules ou contenus ?

L'introspection⁴ est le moyen particulier qui nous permet de connaître nos propres états mentaux occurrents « en première personne », sans passer par l'observation ou l'interprétation du comportement. Les produits de l'introspection sont typiquement des métareprésentations, des représentations de second ordre dont le contenu comprend

-
- 1 Putnam H., *Raison, vérité et histoire*, trad. fr. A. Gerschenfeld, Paris, éditions de Minuit, 1986. Je m'intéresserai surtout à la dernière section du premier chapitre.
 - 2 McDowell J., « Putnam on Mind and Meaning », in *Philosophical Topics*, vol. 20, n°1, 1992, p. 35–48.
 - 3 Voir Putnam H., « Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind », in *The Journal of Philosophy*, vol. 91, n°9, 1994, p. 445–517 et Putnam H., *The Threefold Cord: Mind, Body, and World*, New York, Columbia University Press, 1999.
 - 4 Cette caractérisation liminaire, essentiellement négative, ne présuppose nullement un modèle observationnel de l'introspection. Elle n'exclut pas non plus l'existence d'autres moyens d'acquisition de la connaissance de soi (parfois désignés par le terme « introspection » en un sens plus large). Les positions minoritaires qui nient l'existence de l'introspection ainsi caractérisée débordent le cadre du présent article. Pour un survol de la littérature sur l'introspection, voir Ludwig P. et Michel M., « Introspection », version académique, in M. Kristanek (dir.), *L'Encyclopédie philosophique*, 2017. Je remercie H. Wagner de m'avoir encouragé à clarifier ces points.

une représentation de premier ordre, ainsi qu'une attitude psychologique du sujet envers ce contenu. Par exemple, je peux former sur la base de l'introspection le *jugement* [représentation de second ordre] que je *crois* [attitude] qu'il *pleut* [représentation de premier ordre].

Le débat entre Putnam et McDowell ne concerne pas la nature exacte du moyen par lequel nous arrivons à de telles métareprésentations, c'est-à-dire la méthode de l'introspection, mais plutôt ses cibles ou *objets*. La question n'est ainsi donc pas de savoir comment opère l'introspection, mais plutôt de savoir ce à quoi elle nous donne accès.

La réponse peut paraître triviale, car contenue dans la définition liminaire de l'introspection que nous venons de donner. Par définition, celle-ci permet d'accéder à nos états mentaux occurrents, c'est-à-dire (notamment) à des *représentations mentales* de premier ordre. Si Putnam et McDowell s'accordent sur ce point, un désaccord significatif se fait jour, comme le fait remarquer McDowell⁵ lorsque l'on tient compte de l'ambiguïté de la notion de « représentation ». En effet, ce terme prend des sens très différents chez ces deux auteurs.

« Représentation » peut renvoyer aussi bien au *contenu* – c'est-à-dire à ce qui est représenté, qu'au *véhicule* – c'est-à-dire à ce qui représente. Plus précisément, le contenu d'une représentation, ce sont ses propriétés intentionnelles : ce qu'elle signifie, ou ce à quoi elle fait référence. Les représentations, au sens de « véhicules », ont des natures intrinsèques ou des essences (c'est-à-dire des conditions d'individuation métaphysiques) indépendantes de leurs contenus. Prenez par exemple une occurrence [*token*] du mot « chat ». Il a la propriété de faire référence aux chats. Au-delà de ses propriétés référentielles, il a éventuellement d'autres propriétés intentionnelles, comme par exemple celle d'exprimer un certain sens frégéen. Mais en tant que véhicule, il a aussi d'autres sortes de propriétés. Il a notamment des propriétés *formelles*, qui le caractérisent indépendamment de ce qu'il désigne ou signifie : des propriétés syntaxiques, orthographiques, fonctionnelles, physiques, etc. Par exemple, si vous lisez cet article sur un ordinateur, le mot en question est réalisé physiquement par une certaine configuration de pixels sur votre écran, et a donc une localisation spatio-temporelle. Le mot a aussi des propriétés plus abstraites, comme celle de contenir une occurrence de la lettre-type « a ». Même si l'on perçoit d'ordinaire immédiatement un mot que l'on connaît comme porteur d'une certaine signification, ses propriétés formelles permettent en droit de l'identifier peu importe sa signification.

5 McDowell J., *op. cit.*, p. 43.

En première approche, le désaccord entre Putnam et McDowell est alors le suivant. Selon Putnam, les objets de l'introspection sont des *véhicules* – des représentations mentales au sens de « symboles » mentaux, analogues à des mots ou à des images. Ainsi, Putnam écrit :

« L'introspection ne nous livre pas des “concepts” qui se promènent en tant que tels dans notre esprit. Que l'on interrompe le cours de la pensée, n'importe où, n'importe quand, et l'on percevra des mots, des images, des sensations, des sentiments⁶. »

McDowell interprète la thèse ici défendue par Putnam de la manière suivante : Putnam nierait que l'on puisse accéder par introspection à nos concepts ou à nos pensées – où par « pensées » et « concepts » il faudrait entendre les *contenus* de nos attitudes et leurs constituants.

Or McDowell s'oppose fermement à cette thèse. Selon lui, l'évidence phénoménologique révèle que l'introspection nous livre bien les concepts et pensées (au sens de « contenus ») eux-mêmes :

« Du point de vue d'une phénoménologie libre de préjugés, il serait plus exact de dire que les contenus de la conscience, quand nous avons des pensées occurrentes, sont des pensées elles-mêmes, à peu près à la manière dont Frege emploie le terme “pensée” (ou “Gedanke”) : des sens potentiellement exprimés par des phrases assertoriques, et non pas des véhicules pour de tels sens⁷. »

Un sens frégéen est une entité⁸ abstraite dont la nature intrinsèque est entièrement sémantique – le contenu d'une attitude ou l'un de ses constituants. Si plusieurs sens peuvent avoir le même référent (comme dans le cas des noms propres « Hespérus » et « Phosphorus »), tout sens détermine (à lui seul) un référent ou une extension, indépendamment de l'environnement. Une pensée (ou « Pensée ») frégéenne correspond au sens exprimé par un énoncé déclaratif complet, ou au contenu d'un jugement assertorique. Pour McDowell, les objets de l'introspection sont de telles entités, plutôt que des représentations, c'est-à-dire plutôt que des entités susceptibles d'exprimer des sens, et éventuellement d'exprimer des sens différents en fonction du contexte.

6 Putnam H., *op. cit.*, p. 28.

7 McDowell J., *op. cit.*, p. 44.

8 J'entends simplement par « entité » tout ce à quoi on peut attribuer des propriétés. Par exemple, des « aptitudes » sont des entités en ce sens large. Je remercie un évaluateur anonyme de m'avoir invité à clarifier ce point.

L'assimilation des objets de l'introspection à des *sens fréguéens* en particulier est porteuse de présupposés théoriques lourds : un philosophe qui rejetterait l'existence de telles entités (par exemple un néo-russellien) ne s'y retrouverait pas⁹. Mais on peut formuler de manière plus neutre l'intuition centrale sur laquelle s'appuie l'objection de McDowell contre Putnam, c'est-à-dire d'une manière qui ne se prononce pas sur la nature exacte des contenus de pensée. Il paraît tout simplement faux de dire que l'on accède par introspection à ce *avec quoi* l'on pense, plutôt qu'à ce *à quoi* l'on pense. Ainsi, note McDowell, quand je pense que j'entends le bruit de l'eau qui goutte, je saisis introspectivement que je pense *à de l'eau* ; et quand je visualise le visage de ma femme, je saisis introspectivement qu'il s'agit du visage *de ma femme*¹⁰. J'ai donc accès aux propriétés intentionnelles, et non pas formelles, de mes états mentaux.

Le désaccord entre Putnam et McDowell se réduit-il à un simple conflit d'intuitions phénoménologiques ? Si c'était le cas, il serait sinon inintéressant, du moins difficile à trancher. L'un s'introspecte, et trouve des véhicules ; l'autre fait de même, et trouve des contenus. Peut-être après tout que leurs expériences introspectives respectives sont différentes, et que chacun d'entre eux décrit adéquatement la sienne ?

Présenter ainsi le débat serait en manquer à la fois la nature et l'enjeu. La nature d'abord, parce que ni Putnam, ni McDowell ne se contentent de s'appuyer sur l'introspection comme source d'évidence intuitive. Chacun présente des arguments à propos de l'introspection, que l'on peut évaluer. L'enjeu ensuite, parce que derrière ce débat relativement local concernant les objets de l'introspection, il y a une opposition plus fondamentale entre deux conceptions générales de l'esprit et de la pensée.

Selon McDowell, l'enjeu de la question des objets de l'introspection est en effet de taille. Il s'agirait d'un présupposé majeur et ininterrogé non seulement de la forme modérée d'externalisme que défend Putnam, mais aussi du « problème de la référence », qui traverse toute la philosophie analytique et donne son titre au second chapitre de RVH. Si on abandonne ce présupposé, alors – d'après McDowell encore – on aboutit à une forme plus radicale d'externalisme, et le (pseudo-)problème de la référence est dissous. Quel est donc ce présupposé ?

9 Frege lui-même ne semble pas considérer les sens (les pensées) comme introspectibles indépendamment de leurs véhicules : « La pensée, en elle-même inaccessible au sens, revêt l'habit sensible de la proposition [c'est-à-dire, d'une phrase] et devient ainsi plus saisissable. Nous disons que la proposition exprime une pensée. La pensée échappe aux sens, et tout ce qui est l'objet d'une perception sensible est à exclure du domaine de ce dont on peut examiner la vérité. » dans Frege G., *Écrits logiques et philosophiques*, trad. fr. C. Imbert, Paris, Seuil, 1971, p. 173.

10 McDowell J., *op.cit.*, p. 44.

2. L'enjeu selon McDowell : la théorie représentationnelle de la pensée

Dans RVH, Putnam accepte (implicitement) la *théorie représentationnelle de la pensée*. Selon cette théorie, le fait qu'un sujet *S* a une certaine attitude propositionnelle (par exemple, que *S* croit que *p*) est un fait composite, au sens où la relation attitudinale de *S* à *p* résulte de la composition de deux relations distinctes : d'une part, une certaine relation entre *S* et le véhicule représentationnel de sa pensée ; et d'autre part, le fait que ce véhicule est porteur d'un certain contenu (par exemple, une proposition). En d'autres termes, d'après la théorie représentationnelle, penser, c'est manipuler des symboles – des entités qui ont un contenu, mais dont la nature ne se réduit pas à cela.

La théorie représentationnelle fournit un cadre très général en philosophie de l'esprit, une sorte de schéma théorique qui peut être élaboré de différentes manières. Ce schéma est compatible avec diverses façons de concevoir ses différentes composantes. Ainsi, différentes versions de la théorie envisagent différemment la nature des représentations mentales, le type de relation que le sujet entretient avec elles, ou encore la nature et le mode de détermination des contenus mentaux.

En ce qui concerne ce dernier point, Putnam est justement célèbre pour avoir défendu, notamment grâce à l'expérience de pensée bien connue de Terre-Jumelle, la thèse *externaliste* selon laquelle le contenu représentationnel est déterminé par des relations causales à l'environnement¹¹. Si Putnam applique d'abord cette thèse au contenu linguistique, il en vient à l'étendre au contenu de la pensée¹².

L'externalisme de Putnam demeure cependant *modéré*, dans la mesure où il adopte une conception de la pensée dite « duplex »¹³, c'est-à-dire qui permet sa décomposition en « deux facteurs »¹⁴. Le premier facteur est interne ou « étroit », c'est-à-dire partagé par des jumeaux sur Terre et sur Terre-jumelle. Le second facteur est externe ou « large », et correspond à la différence de contenu référentiel entre leurs pensées. Un défi majeur pour une telle théorie consiste à clarifier la nature du premier facteur, censé être « étroit ». La conception la plus courante est qu'il est soit à identifier avec le véhicule des

11 Putnam H., « The Meaning of "Meaning" », in *Minnesota Studies in the Philosophy of Science*, vol. 7, Minneapolis, University of Minnesota Press, 1975, p. 131-193.

12 Voir aussi Burge T., « Individualism and the Mental », in *Midwest Studies in Philosophy*, vol. 4, n°1, 1979, p. 73-122.

13 McDowell reprend ce qualificatif à McGinn C., « The Structure of Content » (in Andrew Woodfield (dir.), *Thought and Object*, Oxford, Oxford University Press, 1982, p. 207-258) et l'applique à la position de Putnam.

14 Block N., « Advertisement for a Semantics for Psychology », in *Midwest Studies in Philosophy*, vol. 10, n°1, 1986, p. 615-78.

pensées, soit – si on en fait plutôt une forme de *contenu* – déterminé par ce véhicule indépendamment de l'environnement externe. La conception duplex est donc sinon impliquée, du moins fortement suggérée, par la théorie représentationnelle¹⁵.

McDowell s'oppose fondamentalement à Putnam en rejetant la théorie représentationnelle, et la conception duplex. À la place, McDowell endosse un externalisme *radical*, par contraste avec l'externalisme *modéré* de Putnam. Sa position est radicalement externaliste en ce qu'elle refuse non seulement l'idée d'un *contenu* mental étroit, mais même l'idée d'un *véhicule* interne, dont le contenu ou la référence seraient déterminés par des relations externes à l'environnement.

Est-ce à dire que McDowell rejette l'idée que penser, ce soit représenter ? Refuserait-il de dire d'un état mental que c'est une « représentation mentale » ? Seulement si on identifie l'*acte* de représenter, c'est-à-dire *la* représentation en tant qu'exercice d'une capacité, au fait d'avoir des représentations – c'est-à-dire au fait de manipuler des véhicules. La distinction entre les deux est selon McDowell cruciale, et ignorée par Putnam. Comme l'écrit McDowell :

« La possibilité qui disparaît de l'argument de Putnam pourrait être décrite comme la possibilité de la représentation mentale sans représentations [*mental representing without representations*]¹⁶. »

Selon l'externalisme radical de McDowell, penser c'est simplement être en relation avec un *contenu* intentionnel, sans que cette relation soit médiatisée par des entités internes au sujet ou à son esprit, porteuses de ce contenu de manière extrinsèque. Il n'y a donc pas lieu de décomposer nos attitudes en des facteurs internes et externes. Il n'y a pas de véritable pensée¹⁷ sans contenu, et il n'y a de contenu que large.

L'externalisme radical pourrait sembler aller à l'encontre de l'évidence empirique

15 Néanmoins, contrairement à ce que suggère McDowell, la théorie représentationnelle demeure logiquement indépendante de la théorie duplex. Ainsi, on peut accepter une conception externaliste des *véhicules* et donc rejeter la théorie duplex tout en acceptant la théorie représentationnelle. Et on peut accepter une version de la théorie duplex selon laquelle le facteur étroit est un contenu sans véhicule, et donc rejeter la théorie représentationnelle.

16 McDowell J., *op.cit.*, p. 44.

17 La thèse qu'une pensée *véritable* soit la saisie d'un contenu intentionnel ne signifie pas que tout acte mental quel qu'il soit nous mette en rapport avec un contenu. Ainsi, McDowell dans « Singular Thought and the Extent of Inner Space » (in J. McDowell & P. Pettit (dir.), *Subject, Thought, and Context*, Oxford, Clarendon Press, 1986, p. 137-168) défend à la suite d'Evans l'idée selon laquelle les termes singuliers vides expriment seulement des simulacres de pensée [*mock thoughts*]. Je peux selon cette thèse avoir l'*illusion* que j'ai une pensée (que je saisis un contenu intentionnel), mais puisque ma pensée n'a pas de référence, elle n'a pas de contenu. Voir Evans G., *The Varieties of Reference*, Oxford, Oxford University Press, 1982.

selon laquelle notre faculté de penser dépend de l'activité de notre cerveau. Mais ce serait mal l'interpréter. McDowell insiste certes pour rejeter toute conception de l'esprit comme « organe¹⁸ ». Mais l'externaliste radical peut tout à fait admettre qu'il est nécessaire d'avoir un cerveau pour penser. Cependant, il maintiendra que les antécédents causaux de la pensée, ou ses soubassements subpersonnels, n'en sont pas des éléments constitutifs. Selon cette perspective, les états du cerveau ne comptent pas parmi les constituants ontologiques des actes de penser, mais en sont de simples conditions de possibilité nomologique. Leur participation à de tels actes demeure métaphysiquement contingente. Il y a donc une différence nette de nature entre la représentation mentale, qui se passe de véhicules, et la représentation linguistique, qui en implique constitutivement. Ainsi, un acte d'assertion consiste à énoncer une phrase. Mais un acte de jugement ne consiste pas à avoir une phrase – ou tout autre symbole – dans sa tête.

Si Putnam ne va pas aussi loin que McDowell dans son externalisme, est-ce simplement qu'il aurait omis de remarquer que l'on pouvait se passer de la théorie représentationnelle ? Ou a-t-il des raisons d'adopter celle-ci, plutôt que l'externalisme radical ? D'après McDowell, Putnam nierait qu'il se contente de présupposer la théorie représentationnelle, puisque « ce que prétend [Putnam] est que cette thèse est établie par introspection¹⁹ ».

En d'autres termes, du point de vue de McDowell, la dialectique du débat plus global qui l'oppose à Putnam est la suivante. Putnam accepterait la théorie représentationnelle précisément parce qu'il croit que l'introspection nous donne accès aux véhicules de notre pensée. Or, comme on l'a vu, McDowell soutient que Putnam a tort sur ce point : les objets de l'introspection sont des contenus, pas des véhicules. Donc, il n'y a pas de bonne raison de postuler des représentations mentales qui médiatisent notre relation à ce que nous pensons. Et rien n'empêche d'adopter l'externalisme plus radical de McDowell, qui en fait l'économie.

Dans PMM, McDowell ne donne pas vraiment d'arguments *positifs* directement en faveur de l'externalisme radical²⁰. Mais il ne se contente pas non plus de chercher à saper le fondement de la théorie représentationnelle en montrant qu'elle repose sur une

18 McDowell J., PMM, p. 39.

19 *Ibid.*, p. 44.

20 On trouvera de tels arguments par exemple dans Sedivy S., « Minds: Contents Without Vehicles », in *Philosophical Psychology* vol. 17, n°2, 2004, p. 149-181. Des positions différentes, mais de la même famille (contenu sans véhicules) sont défendues par Dennett D., *The Intentional Stance*, Cambridge, MIT Press, 1987, ou encore Stalnaker R., *Inquiry*, Cambridge, MIT Press, 1984.

fausse conception des objets de l'introspection. En effet, McDowell cherche aussi à mettre en avant les effets bénéfiques supposés du rejet de la théorie représentationnelle. L'abandonner permettrait ainsi de dissoudre des problèmes philosophiques majeurs qui la présupposent.

L'un de ceux-ci concerne le problème de la référence mentale. On peut formuler ce problème de la manière suivante : Considérons un état mental m qui est dans la relation de référence R à un objet o . Qu'est-ce qui explique ce type de fait ? C'est-à-dire, qu'est-ce qui explique que m réfère à o plutôt qu'à un autre objet o' ou à rien du tout ?

Pour répondre, il faut trouver une relation (possiblement complexe) R' entre m et o telle que Rmo en vertu du fait que $R'mo$ – c'est-à-dire que l'on cherche le fondement métaphysique²¹ des faits de référence. La relation R' en question doit expliquer la relation R . Pour cela, on doit pouvoir la définir sans présupposer la notion de référence elle-même, sous peine de circularité. Quelles notions présupposent celle de référence mentale ? La réponse n'est pas évidente, mais on peut supposer que R doit être définissable en termes entièrement *non-intentionnels*. En effet, on peut estimer que la référence mentale est la relation intentionnelle la plus fondamentale, originelle, dont les autres dérivent. Si c'est le cas, toute autre notion intentionnelle la présupposera implicitement, et ne peut donc figurer sans circularité vicieuse dans son explication.

En quoi est-ce que ce problème de la référence mentale présupposerait la théorie représentationnelle ? Le raisonnement de McDowell²², reformulé en nos termes, semble être le suivant. Si l'on demande une explication du fait que m réfère à o , c'est que l'on suppose que faire référence à o n'est pas une propriété intrinsèque de m . Or, qu'est-ce qu'un véhicule représentationnel, sinon une entité qui a une propriété intentionnelle (par exemple, celle de faire référence à un objet), mais qui n'a pas cette propriété intrinsèquement ? En d'autres termes, le problème de la référence présuppose que les états mentaux ont une nature intrinsèque qui est autre que purement intentionnelle. Ce qui revient à présupposer que les états mentaux sont, ou impliquent constitutivement, des véhicules représentationnels tels que nous les avons définis à la suite de McDowell.

Mais supposons qu'on accepte plutôt l'externalisme radical. Alors un état mental a son contenu intentionnel intrinsèquement, par nature. Il n'y a donc pas lieu de chercher à expliquer pourquoi ou comment il est venu à l'avoir. On pourrait peut-être même

21 Au sens de Bliss & Trogdon dans Bliss R. & Trogdon K., « Metaphysical Grounding », *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), en ligne (<https://plato.stanford.edu/archives/win2016/entries/grounding>).

22 McDowell J., PMM, p. 44-45.

formuler les choses ainsi : d'après l'externalisme radical, un état mental (une pensée) n'a pas un contenu, il est ce contenu. Il n'y a alors pas lieu d'expliquer par un fait plus primitif le fait qu'il est ce qu'il est, par nature. Ainsi, adopter l'externalisme radical nous permettrait d'éviter de poser ce qui se révélerait être une pseudo-question : « Comment la pensée accroche-t-elle le monde [*hook onto the world*] ? » En d'autres termes, si on rejette la théorie représentationnelle de la pensée, on dissout l'épineux problème de la référence. Ou comme l'écrit McDowell :

« Si l'on prend comme point de départ une conception de la pensée comme en elle-même sans portée référentielle sur le monde, il nous semblera être confrontés à une tâche authentique et urgente, consistant à réintégrer dans notre modèle la manière dont la pensée est dirigée vers le monde. Mais si on n'accepte pas le présupposé selon lequel l'acte de penser, considéré en lui-même, consiste en une manipulation mentale de représentations au sens de Putnam, nous ne sommes confrontés à aucune tâche de cette sorte. Le besoin de construire un "grappin" théorique reliant la pensée au monde ne se fait pas ressentir [...]»²³. »

Le fameux « grappin théorique » [*theoretical hook*] dont parle McDowell, c'est *R'*, la relation non-intentionnelle qui fonderait celle de référence mentale. La quête d'une telle relation, nous dit McDowell, est vaine. Seul le présupposé que penser serait manipuler des représentations symboliques nous ferait miroiter son existence. Et ce présupposé reposerait à son tour sur une mauvaise conception des objets de l'introspection – qui elle-même serait le résultat d'une incapacité à rapporter fidèlement les contenus de cette faculté.

En effet, d'après McDowell, l'introspection ne suggère l'existence de véhicules mentaux qu'à celui qui accepte déjà la théorie représentationnelle, et s'attend donc à les y trouver :

« [Putnam] ne nous dit pas ce qu'il trouve dans son flux de conscience, mais ce qui doit y être, étant donnée la théorie préconçue selon laquelle les contenus de la conscience représentante sont des représentations en son sens²⁴. »

En d'autres termes, selon McDowell, Putnam accepterait la théorie représentationnelle sur la base de l'introspection : Putnam considérerait que celle-ci nous donne accès aux

23 *Ibid.*, p. 45.

24 *Ibid.*, p. 44.

véhicules de nos pensées. Mais sa conception des objets de l'introspection ne serait pas neutre, mais viciée par le fait que Putnam présuppose la théorie représentationnelle de la pensée. La théorie représentationnelle de la pensée ne reposerait donc sur aucun fondement, puisque les données introspectives en sa faveur, bien loin d'être indépendantes de la théorie qu'elles sont censées appuyer, seraient en fait contaminées par l'acceptation préalable de la théorie en question.

3. La position de Putnam clarifiée : la distinction entre concepts et objets mentaux

L'argument que McDowell attribue à Putnam, formulé grossièrement, est le suivant : lorsque je m'introspecte, j'accède à des véhicules représentationnels ; donc, lorsque je pense, je pense avec des véhicules représentationnels. La prémisse est la conception putnamienne des objets de l'introspection. La conclusion est la théorie représentationnelle de la pensée. McDowell objecte que la prémisse présuppose tacitement la conclusion, et que l'argument se révèle à l'examen vicieusement circulaire.

Cependant, McDowell a-t-il raison d'attribuer un tel argument à Putnam ? Il est vrai que l'on trouve parfois un argument visant à établir la théorie représentationnelle sur la base de l'introspection dans la littérature philosophique classique, et même contemporaine²⁵. Mais on ne trouve pas trace d'un tel argument au premier chapitre de RVH²⁶.

Il y a deux excellentes raisons à cela. La première est que le but qui est poursuivi par Putnam dans ce chapitre n'est nullement de fournir un argument *pour* la théorie représentationnelle, mais au contraire d'objecter *contre* certaines versions de celle-ci. La seconde est qu'un tel argument n'aurait d'attrait *que* pour les partisans de versions de la théorie représentationnelle que Putnam s'évertue à critiquer. Ainsi, bien loin de présupposer la version de la théorie représentationnelle qu'endosse Putnam, un tel argument est tout simplement incompatible avec elle : dans le cadre théorique qu'accepte Putnam, l'argument est évidemment fallacieux. En extrayant la thèse de

25 Par exemple, selon P. Carruthers, dans *Language, Thought and Consciousness: An Essay in Philosophical Psychology* (Cambridge, Cambridge University Press, 1998), la théorie qui identifie le langage de la pensée avec le langage naturel fournirait la meilleure explication de l'expérience introspective de la parole intérieure. On remarquera qu'établir la nature linguistique des véhicules mentaux par abduction sur la base de l'introspection ne revient pas tout à fait à l'établir simplement introspectivement. De fait, l'argument de Carruthers, s'il est à mon avis critiquable (pour des raisons que donne E. Machery dans « You Don't Know How You Think: Introspection and Language of Thought », in *British Journal for the Philosophy of Science*, vol. 56, n°3, 2005, p. 469-485) échappe néanmoins aux objections les plus évidentes formulées par McDowell à l'encontre de la position semblable qu'il attribue, selon moi à tort, à Putnam.

26 Ni à ma connaissance ailleurs dans l'œuvre de Putnam dans la période où il accepte la théorie représentationnelle de la pensée.

Putnam sur les objets de l'introspection de son contexte argumentatif d'origine, McDowell en vient à présenter de manière inexacte ce qu'elle vise à établir, et à sous-estimer les arguments en sa faveur.

Resituons donc cette thèse dans son contexte. Putnam aborde la thématique des objets de l'introspection dans le contexte d'une critique de ce qu'il appelle la « théorie magique de la référence », que le premier chapitre dans son ensemble vise à réfuter. Une théorie de la référence est dite « magique » si elle postule que les représentations désignent leurs référents de manière nécessaire ou intrinsèque²⁷. Au contraire, insiste Putnam, la référence est une relation externe, et les rapports entre les représentations et leurs référents sont toujours contingents. Cela est vrai, soutient Putnam, à la fois pour les représentations publiques (« physiques »), et pour les représentations mentales²⁸. Dans le passage qui nous intéresse, Putnam cherche à montrer que cela est vrai *même pour les concepts*.

Pour montrer cela, il lui faut clarifier la nature de ce qu'il appelle « concepts »²⁹. Ceci passe par leur distinction d'avec une autre catégorie d'entités que Putnam appelle « objets mentaux » ou « présentations mentales ». Cette dernière catégorie correspond, selon Putnam, à celle des objets de l'introspection. Elle comprend « des mots, des images, des sensations, des sentiments³⁰ » tels que nous en faisons subjectivement l'expérience dans nos épisodes d'imagerie verbale ou visuelle. Ce qui caractérise les versions de la théorie représentationnelle ciblées par la critique de Putnam, c'est qu'en identifiant – abusivement selon lui – les concepts aux objets mentaux, elles en viennent à être tentées par la théorie magique de la référence.

Reconstruisons – en nos propres termes – la trame argumentative suivie par Putnam. La théorie magique de la référence est fautive des véhicules presque par définition. En effet, on a défini ceux-ci comme des entités qui ont des propriétés intrinsèques qui ne sont *pas* intentionnelles ou référentielles. Il s'ensuit que le fait d'avoir telle ou telle signification ou de désigner tel ou tel référent ne fait pas partie de leur nature propre.

27 Putnam H., RVH, p. 23.

28 *Ibid.*, p. 13.

29 Ce terme, comme celui de « représentation », et de manière parallèle, n'a pas la même signification chez Putnam et chez McDowell. Il s'agit, tel que j'interprète Putnam, des représentations mentales qui sont les véhicules porteurs des contenus de nos pensées, et qui sont individuées par leur rôle fonctionnel/inférentiel étroit. La notion hérite de l'ambiguïté de celle de « représentation ». (Comme me l'a fait remarquer H. Wagner, cette interprétation, qui me semble rendre compte de manière charitable de ce que dit Putnam dans RVH, est potentiellement en tension avec une lecture plus « wittgensteinienne » qui identifierait les concepts avec des « aptitudes ». Si tension il y a, elle me semble être interne à la pensée de Putnam.)

30 *Ibid.*, p. 28.

C'est en fait un peu rapide – d'où le « presque ». Un véhicule représentationnel a intrinsèquement des propriétés non-intentionnelles, certes. Mais cela n'exclut pas d'emblée qu'il ait *aussi* des propriétés intentionnelles de manière nécessaire. Pourquoi la même entité ne pourrait-elle pas avoir essentiellement ces deux types de propriétés à la fois ?

En fait, ce que cherche à montrer Putnam est plutôt que les propriétés intentionnelles d'une représentation, et en particulier ses propriétés référentielles, ne découlent jamais de manière nécessaire de ses propriétés intrinsèques non-intentionnelles. Cela semble indéniable quand les propriétés non-intentionnelles de la représentation qu'on considère sont simplement ses propriétés physiques ou formelles. Ainsi, Putnam a souligné au début du premier chapitre que la ressemblance objective entre une figure tracée par une fourmi dans le sable et Churchill ne fait pas de la figure un portrait *de* Churchill, c'est-à-dire lui faisant référence³¹.

Mais la fausseté de la théorie magique est moins évidente si on considère une autre variété de propriétés possédées par certaines représentations mentales – à savoir les propriétés *phénoménales*. Les états paradigmatiques qui ont ce type de propriétés sont les états *sensoriels* ou *perceptuels* conscients – par exemple, une sensation de douleur aiguë, ou une expérience visuelle d'une certaine teinte de rouge. Le *caractère phénoménal* d'un état mental est « ce que cela fait »³² du point de vue subjectif d'être dans cet état.

Reconsidérons à présent le type d'entités que Putnam décrit comme des « objets mentaux » ou « présentations mentales ». McDowell explique³³ qu'il s'agit selon Putnam de *véhicules* (« des éléments qui ne sont pas intrinsèquement représentationnels »). C'est vrai, si on entend par là que ce sont des entités dont les conditions d'individuation ne sont pas leurs propriétés référentielles ou intentionnelles. Mais il ne s'agit certainement pas d'une liste d'entités mentales dont les conditions d'individuation seraient physiques ou formelles, comme les mots du langage public, ou comme les représentations mentales, tels que les conçoit la théorie computationnelle-représentationnelle de la pensée³⁴. En effet, il est frappant qu'il s'agisse d'entités caractérisées par leurs propriétés *phénoménales*.

Il faut ainsi distinguer deux conceptions très différentes des véhicules mentaux, en

31 *Ibid.*, p. 11.

32 Nagel T., « What Is It Like to Be A Bat? », in *Philosophical Review*, vol. 83, 1974, p. 435-50.

33 Putnam H., RVH, p. 44.

34 Par exemple chez Fodor J., *The Language of Thought*, Cambridge, Harvard University Press, 1975.

fonction du type de propriétés non-intentionnelles ou non-référentielles qu'on leur prête³⁵. Dans la tradition *fonctionnaliste* à laquelle appartient Putnam³⁶, les véhicules mentaux sont individués soit par leurs propriétés physiques, soit par leurs rôles fonctionnels, computationnels ou inférentiels³⁷ – c'est-à-dire (en gros) ce que Putnam, dans une tradition wittgensteinienne (ou sellarsienne), décrit comme leur « emploi³⁸ ». Mais dans les traditions *empiristes* et *phénoménologiques*, les véhicules mentaux possèdent essentiellement, et sont individués par, leurs caractères *sensoriels* ou *phénoménaux*.

Même si l'on est prêt à admettre que les propriétés formelles/fonctionnelles d'une représentation ne déterminent pas sa référence, il est moins évident de séparer son caractère phénoménal de son contenu intentionnel. On peut ainsi être tenté par la thèse d'inspiration phénoménologique selon laquelle le caractère phénoménal de la pensée détermine son contenu intentionnel³⁹.

C'est précisément ce genre de position que Putnam attaque dans la section qui constitue le contexte immédiat dans lequel est avancée la thèse selon laquelle les objets de l'introspection sont des objets mentaux, et non des concepts. Plus spécifiquement, les cibles de Putnam (tel que je le comprends) sont les versions de la théorie représentationnelle opposées à la sienne, et qui soutiennent que :

i) Les concepts (véhicules de nos pensées) ont des caractères phénoménaux/sensoriels de manière essentielle – comme les idées des empiristes classiques ou contemporains⁴⁰.

35 Sedivy S., *op.cit.*

36 Putnam H., *Philosophical Papers, Volume 2: Mind, Language and Reality*, Cambridge, Cambridge University Press, 1979.

37 On considère souvent que ces rôles dérivent des rôles causaux des représentations.

38 Putnam H., RVH, p. 29.

39 Ainsi, d'après certains philosophes contemporains, dont certains tirent inspiration de Brentano et de la tradition phénoménologique, le caractère phénoménal intrinsèque de nos concepts détermine leurs *conditions* de référence, qui peuvent être satisfaites par différents objets dans différents contextes d'évaluation (Farkas K., « Phenomenal Intentionality without Compromise », in *The Monist*, vol. 91, n°2, 2008, p. 273-93). Selon la thèse dite de « l'intentionnalité phénoménale », aujourd'hui largement défendue, le caractère phénoménal est même *la* source de l'intentionnalité non-dérivée des états mentaux en général (voir Bourget D. & Mendelovici A., « Phenomenal Intentionality », *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.), en ligne (<https://plato.stanford.edu/archives/spr2019/entries/phenomenal-intentionality>)).

40 On peut accepter cette thèse empiriste sans accepter que les propriétés sensorielles d'une représentation déterminent sa référence. Voir par exemple Prinz J., *Furnishing the Mind: Concepts and Their Perceptual Basis*, Cambridge, MIT Press, 2002.

ii) Les caractères phénoménaux/sensoriels des concepts déterminent leur contenu intentionnel ou référentiel – comme chez Brentano (que Putnam cible explicitement⁴¹) et de nombreux phénoménologues.

D'après Putnam, l'erreur de telles positions est double. D'une part, contre i), elles confondent les concepts et les objets mentaux/présentations mentales. Seule la seconde catégorie est individuée par ses propriétés phénoménales, et est accessible à l'introspection. D'autre part, contre ii), le caractère phénoménaux d'une représentation ne détermine pas son contenu ou sa référence. Il s'ensuit que la théorie magique de la référence est fautive même pour les concepts, puisque les concepts ne sont pas individuels par leurs propriétés phénoménales, et que même s'ils l'étaient, ces propriétés ne déterminent de toute manière pas la référence.

Nous reviendrons aux arguments que fournit Putnam pour cette position par la suite. Mais notons immédiatement qu'il s'inscrit dans la continuité d'une tradition analytique bien établie de critique des conceptions empiristes ou phénoménologiques des concepts et de la pensée. Putnam souligne ainsi à plusieurs reprises qu'il s'inspire de Wittgenstein⁴², et la parenté avec certains passages anti-psychologistes de Frege⁴³, ou anti-empiristes de Sellars⁴⁴ est patente.

Nous sommes maintenant en mesure d'expliquer clairement pourquoi l'argument que McDowell prête à Putnam est en réalité *incompatible* avec sa position réelle. L'argument serait évidemment fallacieux dans le cadre de la théorie de Putnam car il reposerait sur une ambiguïté. La notion de « véhicule » qui figure dans la prémisse n'est pas la même que celle de « véhicule » qui figure dans la conclusion. Ou, pour le dire autrement, il manque à l'argument une prémisse intermédiaire selon laquelle les véhicules accessibles à l'introspection sont les véhicules qui sont manipulés dans la pensée. Or, Putnam rejette précisément une telle identification. En effet, les premiers correspondent pour lui aux objets mentaux, et les seconds aux concepts. C'est seulement si l'on identifie objets mentaux/présentations mentales, d'une part, et concepts, d'autre part, qu'on pourrait être tenté d'établir la théorie représentationnelle de la pensée sur la base de l'introspection. Puisque Putnam dénonce une telle identification comme une dangereuse confusion menant tout droit à la fallacieuse théorie magique de la référence, il est

41 Putnam H., RVH, p. 27.

42 *Ibid.*, p. 31-32.

43 Frege G., *op.cit.*

44 Sellars W., « Empiricism and the Philosophy of Mind », in *Minnesota Studies in the Philosophy of Science* vol. 1, Minneapolis, University of Minnesota Press, 1956, p. 253-329.

trompeur de suggérer, comme le fait McDowell, qu'il accepterait un tel argument.

L'argument repose sur une confusion entre *imagerie* verbale (visuelle, etc.) et *pensée*⁴⁵. Qu'est-ce donc que la parole intérieure, si elle n'est pas à confondre avec le véhicule de la pensée ? Putnam, dont le but n'est pas de donner une théorie positive de la nature de la parole intérieure, ne se prononce pas très précisément sur la question, aujourd'hui très discutée⁴⁶. Mais ses quelques remarques sur le sujet suggèrent fortement les types de réponse qu'il devrait accepter. Dans le cadre de la théorie représentationnelle de la pensée, il y a au moins deux conceptions plausibles de la nature de la parole intérieure qui opposent nettement celle-ci à la pensée elle-même⁴⁷.

Selon la première conception, la parole intérieure est une forme de parole « subvocale ». Plus précisément, elle consisterait en des actes de parole silencieux avortés ou inhibés à un certain stade de leur production, avant la prononciation effective⁴⁸. Une donnée compatible avec cette hypothèse provient de l'étude des hallucinations verbales chez certains schizophrènes. L'observation que ces hallucinations s'accompagnent parfois de vocalisations ouvre la possibilité que les « voix » que les patients entendent « dans leur tête » correspondent à des paroles inaudibles produites par eux, mais dont ils ne sont pas conscients d'être la source⁴⁹. Cette hypothèse prédit qu'il devrait être possible d'amplifier ces paroles de manière à rendre les « voix » publiquement audibles. Ainsi, en attachant des microphones au larynx d'un patient schizophrène, Green et Preston⁵⁰ ont pu enregistrer des ordres (à la troisième personne) que le patient se chuchotait à lui-même. Le contenu enregistré correspondait précisément à ce que le patient rapportait être des ordres qu'il entendait

45 Notons que cette distinction est orthodoxe dans la tradition fonctionnaliste, notamment chez les auteurs qui défendent la théorie computationnelle-représentationnelle de la pensée. Ainsi, Fodor, *op.cit.*, qui défend la théorie du langage de la pensée nie farouchement que celui-ci soit accessible à l'introspection, et le distingue soigneusement de la parole intérieure.

46 Langland-Hassan P. & Vicente A. (dir.) *Inner Speech: New Voices*, Oxford, Oxford University Press, 2018.

47 Gregory D., *Inner Speech : A Philosophical Analysis*, thèse de doctorat en philosophie sous la direction de Daniel Stoljar, Australian National University, 2017, 152 p.

48 Vicente A. & Martinez-Manrique F., « The Nature of Unsymbolized Thinking », in *Philosophical Explorations*, vol. 19, n°2, 2016, p. 173-187.

49 Ceci est compatible avec l'un des modèles explicatifs de certains troubles schizophréniques, le modèle dit du « comparateur » (Frith C., « The Self in Action: Lessons from Delusions of Control », in *Consciousness and Cognition*, vol. 14, n°4, 2005, p. 752-770), qui postule que ceux-ci proviennent d'une incapacité à distinguer les effets de ses propres actions de ceux de forces extérieures, une tendance qui se manifeste par exemple dans la capacité accrue de certains schizophrènes à se chatouiller eux-mêmes (Blakemore S., Wolpert D. & Frith C., « Central Cancellation of Self-Produced Tickle Sensation », in *Nature Neuroscience*, vol. 1, n°7, 1998, p. 635-640).

50 Green P. & Preston M., « Reinforcement of Vocal Correlates of Auditory Hallucinations by Auditory Feedback: A Case Study », in *The British Journal of Psychiatry*, vol. 139, n°3, 1981, p. 204-208.

une voix (féminine) lui donner, voix qu'il attribuait à une personne réelle distincte de lui-même.

Si les paroles intérieures sont ainsi effectivement un type de *paroles* qui se trouvent ne pas être vocalisées, alors ce sont des véhicules certes, mais *linguistiques* – qui partagent certaines des propriétés (par exemple, phonologiques) des énoncés publics. Ce ne sont pas des véhicules *de pensée*.

Selon la seconde conception, faire l'expérience de paroles intérieures, c'est *imaginer* des entités linguistiques comme des mots ou des paroles, ou imaginer en produire⁵¹. Ainsi, tout comme vous pouvez imaginer un chat, vous pouvez imaginer le mot « chat ». Dans ce cas, vous vous représentez les propriétés de ce mot, sous le mode imaginatif (qu'on peut traiter dans le cadre de la théorie représentationnelle comme une certaine attitude propositionnelle). Parmi ces propriétés représentées, il pourra y avoir les propriétés formelles du mot – par exemple son orthographe ou sa prononciation. Ainsi, le test psychologique habituel⁵² pour déterminer si la faculté de parole intérieure d'un sujet est intacte consiste à lui faire accomplir la tâche suivante : on lui montre des paires de mots visuellement dissimilaires (par exemple, « Adam » et « dent »), et on lui demande de juger en silence (sans les prononcer) s'ils riment. Un tel test incarne l'idée que dans l'imagerie verbale, le sujet *représente* les sons du langage, ou y fait référence. Cela revient à dire que les mots mentaux ne sont pas les *véhicules* de la pensée, mais au contraire des éléments représentés dans son *contenu* – à savoir, le contenu représentationnel de l'attitude d'imagination.

Ainsi, du point de vue de cette conception de la parole intérieure comme une forme d'imagination, si on identifie les mots mentaux et les concepts, on confond ce *à quoi* on pense, d'une part, avec ce *avec quoi* on pense, d'autre part⁵³. Les mots mentaux sont le *contenu* de ce qu'on imagine (l'une de nos attitudes). Les concepts sont quant à eux les *véhicules* de nos attitudes en général. En représentant un mot mental, il nous faut un concept *de mot*, c'est-à-dire y faisant référence. Ce concept ne peut pas lui-même être identique au mot dont il est le concept. Affirmer le contraire serait commettre une erreur de la même famille que la confusion entre usage et mention, ou de ce que Pylyshyn⁵⁴ appelle le « sophisme intentionnel », qui consiste à confondre les propriétés

51 Machery E., *op.cit.*

52 Comme le note Langland-Hassan P., « Inner Speech and Metacognition: In Search of a Connection », in *Mind and Language*, vol. 29, n°5, 2014, p. 511-533.

53 Machery E., *op.cit.*

54 Pylyshyn Z., *Computation and Cognition*, Cambridge, MIT Press, 1984.

d'un véhicule avec celles représentées dans son contenu.

Nul besoin, pour notre propos, de trancher entre les deux conceptions de la parole intérieure que nous venons de résumer brièvement. Il n'est pas clair qu'elles soient mutuellement exclusives (il pourrait s'agir de deux sous-phénomènes différents), ni exhaustives. Ce qui compte est qu'elles sont chacune compatibles avec la distinction putnamienne entre objets mentaux et concepts. Et une fois cette distinction clairement établie, il nous est possible de clarifier la position réelle de Putnam, et ses éventuels présupposés.

Contrairement à ce qu'affirme McDowell, la conception putnamienne des objets de l'introspection comme objets mentaux plutôt que comme concepts ne présuppose pas la théorie représentationnelle – pas plus qu'elle ne fournit d'argument en faveur de cette théorie. Et cela, pour la simple raison que cette conception des objets de l'introspection, si elle est entièrement compatible avec la théorie représentationnelle, est en fait logiquement indépendante de celle-ci.

Une façon de s'en persuader est de noter qu'elle est en fait compatible avec l'externalisme radical à la McDowell. L'externalisme radical nie l'existence de véhicules de pensée. Mais identifier les objets de l'introspection à des objets mentaux, c'est-à-dire à des entités de la catégorie ontologique des paroles intérieures, n'implique nullement, et n'est nullement impliquée par, l'existence de tels véhicules de pensée. En effet, les mots mentaux sont soit des véhicules représentationnels linguistiques de même nature que les paroles ou énoncés publics, soit la représentation mentale de telles représentations.

Supposons qu'on accepte l'externalisme radical. On ne nie certainement pas pour autant l'existence de véhicules *linguistiques*, comme les paroles : McDowell limite précisément sa thèse anti-représentationnaliste à la pensée. Et on ne nie pas non plus que des véhicules linguistiques puissent figurer dans le *contenu* de la pensée ou de l'imagination. Après tout, McDowell admet certainement qu'on puisse imaginer des énoncés – ou même imaginer des représentations mentales, que celles-ci existent ou non. Cela est en effet entièrement compatible avec l'hypothèse externaliste radicale selon laquelle la pensée met le sujet immédiatement en rapport avec le contenu de ce qu'il pense, sans médiatisation par des représentations mentales.

De même, la théorie représentationnelle de la pensée en tant que telle n'implique nullement que les objets de l'introspection soient des objets mentaux – même si ces thèses sont évidemment compatibles. En effet, les objets de l'introspection sont ceux qui

sont représentés dans le *contenu* de l'introspection. Et la théorie représentationnelle est simplement neutre quant à la nature de ceux-ci. La théorie représentationnelle de la pensée affirme que penser, c'est manipuler des représentations ; mais elle ne se prononce pas sur l'introspectibilité de celles-ci.

Toutefois, Putnam, lui, se prononce bien sur la nature des objets de l'introspection. Sa position à ce sujet a deux versants, l'un positif, l'autre négatif. Le versant positif est la thèse selon laquelle les objets mentaux sont accessibles à l'introspection. Nous avons vu que, contrairement à ce que suggère McDowell, cela n'implique nullement que les véhicules *de la pensée* sont introspectibles. En effet, le versant négatif de la position de Putnam est que les *concepts*, eux, ne sont *pas* introspectibles (du moins « en tant que tels⁵⁵ » – nous y reviendrons). McDowell interprète cette thèse comme signifiant que les *contenus* de nos pensées ne sont pas introspectibles. Or, chez Putnam, les « concepts » sont identifiés plutôt à des représentations mentales individuées par leur rôle fonctionnel ou inférentiel étroit⁵⁶, et non à des constituants de contenu (comme des sens frégéens). Donc la non-introspectibilité des concepts au sens de Putnam n'implique pas immédiatement la non-introspectibilité des concepts au sens de McDowell. Néanmoins, le désaccord entre eux n'est pas verbal. Si les concepts au sens de Putnam ne sont pas introspectibles, *a fortiori* le contenu référentiel ou frégéen ne l'est pas non plus. McDowell a donc raison d'attribuer à Putnam la thèse selon laquelle les constituants des contenus de nos pensées ne sont *pas* introspectibles. Et il est vrai qu'une telle thèse, tout comme celle de la non-introspectibilité des concepts *quelle que soit leur nature*, peut paraître – à première vue – hautement contre-intuitive. Putnam serait-il en train d'affirmer que l'introspection ne nous permet pas de savoir ce que l'on pense – c'est-à-dire le contenu de nos propres pensées, et les concepts qui s'y déploient ?

4. Les arguments de Putnam

Une fois certaines fausses interprétations écartées, la position de Putnam est moins contre-intuitive qu'il n'y paraît : elle ne s'oppose nullement à l'idée d'un accès privilégié « en première personne » au contenu de nos pensées, par exemple. Afin d'en mieux saisir la teneur réelle, commençons par rappeler les arguments que Putnam donne en sa faveur. Il y en a trois principaux. Deux de ceux-ci sont des expériences de pensée. Le troisième est un argument de principe.

55 Putnam H., RVH, p. 28.

56 Ceci rend compte notamment de ses affirmations (*ibid.*, p. 29) selon lesquelles les jumeaux sur Terre et sur Terre Jumelle ont les mêmes concepts, et selon laquelle il a le même concept d'orme que de bouleau.

Les deux expériences de pensée visent à illustrer deux possibilités converses. Selon Putnam, deux sujets (ou le même à différents moments) peuvent être (ce que j'appellerai) des *doubles introspectifs*, c'est-à-dire avoir accès par l'introspection aux mêmes entités mentales, mais ne pas être des *doubles conceptuels*, c'est-à-dire ne pas exercer (ou posséder) les mêmes concepts. A l'inverse, deux sujets peuvent ne pas être des doubles introspectifs, mais être néanmoins des doubles conceptuels.

Le premier cas correspond à celui où un sujet ferait l'expérience des mêmes épisodes d'imagerie verbale, visuelle, etc. que moi, mais sans avoir les mêmes pensées. Ainsi, tandis que je *comprends* mon monologue intérieur, c'est-à-dire que j'associe à mes paroles mentales les concepts appropriés, mon double introspectif aurait seulement l'*illusion* de comprendre les siennes :

« Je n'ai pas de mal à imaginer quelqu'un qui penserait précisément ces mots-ci, qui se les réciterait, qui aurait vraiment l'impression de les comprendre, de faire des affirmations, et qui, une minute plus tard, en émergeant d'une transe hypnotique, s'apercevrait qu'il ne comprenait pas du tout les mots qui lui passaient par la tête et qu'il ne connaissait même pas la langue dont ils étaient tirés⁵⁷. »

Putnam précise qu'il s'agit d'une simple expérience de pensée, visant seulement à établir la *concevabilité* d'un tel cas. Cependant, il existe des cas empiriques qu'on pourrait interpréter comme analogue à celui-ci, dans la mesure où ils semblent mettre également en jeu une forme d'*illusion introspective* – un cas où un sujet croit sur la base de l'introspection être dans un certain état mental, mais où l'on peut raisonnablement douter qu'il le soit effectivement. Ainsi, un sujet atteint du syndrome d'Anton⁵⁸ juge qu'il est en train d'avoir des expériences perceptuelles plus ou moins normales. Mais si on l'interroge, on se rend compte qu'il n'a pas accès au contenu de ce qu'il croit voir (les objets et propriétés représentés), et donc que le rôle fonctionnel/inférentiel de l'état mental de premier ordre dans lequel il se trouve n'est pas celui correspondant à une perception normale, comme l'illustre l'échange suivant (G.G. est l'expérimentateur, et H.S. le patient) :

« G. G. : (ouvre et ferme une paire de ciseaux) : Avez-vous une idée de ce que cela pourrait être ?

57 *Ibid.*, p. 28.

58 Voir Macpherson F., « A Disjunctive Theory of Introspection: A Reflection on Zombies and Anton's Syndrome », in *Philosophical Issues*, vol. 20, n°1, 2010, p. 226-265.

H. S. : Des ciseaux ?

G. G. : Les voyez-vous ?

H. S. : Seulement vaguement. J'ai un peu deviné.

G. G. (cache silencieusement les ciseaux sous la table) [la partie de la conversation en italiques a lieu pendant qu'ils sont cachés]) :

Que voyez-vous de ces ciseaux ?

H. S. : Vers le haut il y a les poignées par lesquelles on les prend, et en dessous il y a la partie pour couper.

G. G. : *Êtes-vous en train de voir cela ?*

H. S. : *Oui*⁵⁹. »

Ici, le sujet dit qu'il voit des ciseaux que l'expérimentateur a en fait dissimulés sous une table, et il paraît sincère. Ainsi, il croit voir quelque chose qu'il ne voit pas. Bien sûr, on peut interpréter de telles données empiriques de plusieurs manières. Mais une interprétation possible⁶⁰ est que le sujet juge qu'il est dans un état mental avec le caractère phénoménal d'une expérience visuelle, mais qu'il n'est pas dans l'état en question. Il s'agit ici d'un cas où un sujet à l'illusion d'avoir une expérience visuelle, alors que le cas envisagé par Putnam est un cas où un sujet à l'illusion d'exercer des concepts. Pour rendre l'exemple empirique plus similaire à celui envisagé par Putnam, supposons que le sujet en vienne à croire qu'il exerce un concept démonstratif perceptuel, comme CETTE PAIRE DE CISEAUX. Si le déploiement d'un concept de ce type nécessite la possession d'un percept⁶¹ correspondant, c'est-à-dire une expérience visuelle des ciseaux, alors on aurait potentiellement un cas où le sujet a l'illusion d'exercer un concept qu'il ne possède pas.

Le second cas imaginé par Putnam est moins baroque. Il s'agit simplement d'un sujet qui exerce des concepts, entretient des pensées et forme des jugements, mais n'a pas d'imagerie verbale ou visuelle attenante :

59 Goldenberg G., Müllbacher W. & Nowak A., « Imagery Without Perception – A Case Study of Anosognosia for Cortical Blindness », in *Neuropsychologia*, vol. 33, 1995, p. 1373-1382, p. 1378, cité dans Macpherson F., *ibid.*

60 Macpherson F., *ibid.*

61 Voir par exemple la théorie des concepts démonstratifs perceptuels de Bach K., *Thought and Reference*, Oxford, Oxford University Press, 1987.

« Considérons maintenant l'homme parfaitement possible qui n'a aucun "monologue intérieur". Il parle parfaitement l'anglais, et si on lui demande son avis sur quelque chose, il l'expose longuement. Mais quand il ne parle pas à haute voix il ne pense jamais avec des mots, des images et ainsi de suite, et rien ne lui passe par la tête. [...] Il n'a pas de monologue intérieur, mais ses actions sont intelligentes et orientées, et si quelqu'un venait le voir et lui demandait "qu'est-ce que vous faites ?", il donnerait des réponses parfaitement cohérentes⁶². »

Dans un tel cas, nous dit Putnam, il n'y a pas d'obstacle à attribuer au sujet des pensées conscientes et des concepts. Le cas n'est pas si surprenant dans la mesure où nombre de sujets déclarent faire l'expérience de « pensées asymbolisées [*unsymbolized thinking*] », phénomène défini par Hurlburt & Akther (2008) comme « l'expérience d'une pensée explicite, différenciée, qui n'inclut pas l'expérience de mots, d'images, ni d'aucun autre symbole ». Ces psychologues décrivent le phénomène comme « fréquent », quoiqu'il ne soit pas « omniprésent » : dans leur étude empirique, qui s'appuie sur les rapports introspectifs de sujets, les pensées asymbolisées correspondent à 22 % des pensées rapportées, et environ 75 % des sujets rapportent avoir au moins certaines pensées de ce type. Bien sûr, on peut douter de la fiabilité de tels rapports. Mais il paraît difficile de nier la *concevabilité* des pensées asymbolisées – et de la situation qu'imagine Putnam.

Venons-en à l'argument de principe de Putnam pour distinguer les objets de l'introspection des concepts. Celui-ci repose sur l'idée selon laquelle le critère de possession d'un concept est « la capacité d'agir d'une certaine manière ». La compétence pratique en question correspond, pour le dire vite, au rôle fonctionnel/inférentiel caractéristique du concept. Ainsi, quiconque a une image ou un mot mental d'arbre, mais aucune capacité de catégorisation correspondante – aucune capacité à distinguer les percepts d'arbres des non-arbres – n'a pas selon ce critère le concept arbre. Tandis que le sujet qui a au contraire toutes les capacités pertinentes, par exemple qui peut sans difficulté reconnaître des arbres, etc. a le concept arbre, même s'il n'a pas en son for intérieur d'imagerie verbale ou visuelle attenante.

Or, selon Putnam, cette compréhension de la nature des concepts implique qu'ils ne sont pas introspectibles, pour des raisons ontologiques : ce qui est introspectible, ce sont des états mentaux *occurrents*, c'est-à-dire des événements ; mais un concept, souligne

62 Putnam H., RVH, p. 31.

Putnam, est une *capacité* – et donc pas une occurrence⁶³. On pourrait ici songer à une analogie avec la perception visuelle. On ne peut voir que ce qui est présent ou a lieu dans notre champ de vision⁶⁴. Peut-être certaines propriétés dispositionnelles des objets sont-elles visibles – par exemple, la fragilité d'un verre⁶⁵. Mais il semble peu plausible que l'on puisse littéralement voir une propriété dispositionnelle, abstraite, et profondément relationnelle comme une capacité conceptuelle. De même, selon Putnam, tout au plus a-t-on accès par l'introspection à des qualités introspectibles *corrélées* avec les concepts, mais jamais aux capacités constitutives des concepts eux-mêmes⁶⁶. Par exemple, on peut introspecter l'image mentale d'arbre qui accompagne l'exercice du concept orme, mais la même image pourrait accompagner l'image d'un autre concept, comme celui du concept bouleau, ou aucun concept du tout.

Les trois arguments qu'on vient de donner conduisent à distinguer nettement les objets mentaux ou présentations mentales qui accompagnent la pensée, de l'exercice des concepts qui la constituent. Penser n'est pas parler dans sa tête ou y voir des images, mais exercer certaines capacités conceptuelles. Toutefois, ne peut-on accepter une telle distinction, mais refuser d'en tirer la conclusion que les concepts et la pensée eux-mêmes sont inaccessibles à l'introspection, comme semble le soutenir Putnam ? C'est là précisément le point de vue de McDowell, dont il nous faut à présent examiner les objections.

5. Les objections de McDowell

Commençons par la réponse de McDowell à l'argument de principe de Putnam, puisque si cet argument fonctionne, on pourrait supposer que les autres sont sinon superflus, du moins secondaires. McDowell⁶⁷ note à juste titre que même si on identifie la *possession* d'un concept à celle d'une capacité, il est néanmoins naturel d'identifier l'*exercice* d'un concept, c'est-à-dire son déploiement effectif, à un événement occurrent. Or il n'y a pas

63 *Ibid.*, p. 31.

64 Des phénomènes comme la persistance rétinienne ou la perception amodale demanderaient à ce qu'on nuance une telle affirmation.

65 On pourrait objecter qu'on juge que de telles propriétés sont instanciées sur la base de la perception plutôt qu'on ne les voit, à strictement parler.

66 Un lecteur anonyme suggère qu'une critique possible de Putnam serait l'idée qu'on puisse accéder par introspection à un mot mental, conçu comme la base catégorique d'un concept (lui-même dispositionnel), de sorte qu'accéder au mot mental reviendrait à accéder au concept. Il me semble que Putnam rejeterait l'identification des mots mentaux avec même la base catégorique des concepts, identification qui supposerait, comme le note à juste titre ce lecteur, une forme de phénoménologie cognitive.

67 McDowell J., PMM, p. 45, note 24.

d'obstacle ontologique à ce que l'exercice d'un concept soit introspectible. Ni à ce que celui-ci hérite des propriétés essentielles de la capacité dont il est exercice, de sorte que le débat de savoir s'il faut individuer cette capacité de manière large ou étroite reste entier.

En réponse, Putnam pourrait insister qu'ontologiquement, la notion de concept en tant que capacité a une primauté sur celle de concept en tant qu'occurrence, et qu'après tout, il ne soutient pas tant que les concepts ne sont pas introspectibles, mais plutôt que « l'introspection ne nous livre pas des "concepts" qui se promènent *en tant que tels* [c'est nous qui soulignons] dans notre esprit »⁶⁸. En effet, un énoncé de la forme *S introspecte ...* a deux lectures : l'une où ... est un contexte référentiellement transparent, l'autre où ... est un contexte référentiellement opaque. Dans la seconde lecture, on introspecte un objet non pas tout court, mais *en tant que* possesseur de telles ou telles propriétés⁶⁹. La thèse de Putnam pourrait alors s'interpréter comme s'appliquant à cette lecture opaque uniquement : Putnam ne nierait pas que les concepts soient introspectibles tout court, mais qu'ils le soient *en tant que possesseurs des propriétés qui les individuent/constituent en tant que concepts*. Une telle position demeurerait justifiée par l'observation, nullement réfutée par McDowell, selon laquelle les propriétés fonctionnelles/inférentielles permettant d'individuer les concepts ne sont pas de nature à être introspectibles.

Quoi qu'il en soit, l'objection de McDowell demeure globalement convaincante. Il ne suffit pas de faire appel à l'idée que le critère de possession d'un concept est la possession d'une capacité pour établir que les concepts ne sont pas introspectibles. De plus, la thèse de l'absence d'introspectibilité des concepts fait face à une objection sérieuse.

Quand on pense, on sait immédiatement ce que l'on pense, c'est-à-dire à quoi l'on pense. Comme le note McDowell, je n'ai pas besoin d'*interpréter* les objets de l'introspection. Quand je forme une pensée consciente, je n'ai pas à l'esprit un simple véhicule syntaxique dont j'aurais à extraire la signification. Ainsi, quand me vient la

68 Putnam H., RVH, p. 17.

69 D'après la lecture transparente, si $o = o' = le F$, alors *S introspecte o'* (ou *S introspecte le F*) est vrai ssi *S introspecte o* l'est. Mais, d'après la lecture opaque, S peut introspecter *o en tant que o*, mais pas *en tant que o'* ou pas *en tant que F*. Je m'appuie ici sur la distinction d'Anna Giustina & Uriah Kriegel (« Fact-Introspection, Thing-Introspection, and Inner Awareness », in *Review of Philosophy and Psychology*, vol. 8, n°1, 2017, p. 143-164) entre introspection objectuelle (*thing introspection*) et factuelle (*fact introspection*). L'opacité référentielle de l'introspection est aussi mise en avant par P. Churchland dans *Matter and Consciousness*, Cambridge, MIT Press, 1984, qui défend la thèse que l'apprentissage perceptuel pourrait permettre d'introspecter les véhicules cérébraux *en tant que tels*.

pensée que j'entends le son de l'eau qui goutte, je ne suis pas confronté à « quelque chose dans laquelle on peut lire la signification que l'on entend le bruit de l'eau, comme on peut la lire dans la forme du signe "J'entends le bruit de l'eau qui goutte" ». En effet,

« il semble entièrement inexact de suggérer que quand la pensée me frappe que j'entends le bruit de l'eau qui goutte, le fait que ma pensée, est, disons, à propos de *l'eau* ne fait pas partie de ce que je trouve dans mon flux de conscience, mais doit être lu dans ce que j'y trouve »⁷⁰.

Au contraire, ce à quoi l'on pense, et les concepts qu'on déploie, ont un impact caractéristique sur mon expérience subjective, et font donc partie de ce qui m'est introspectible. Ainsi, en ce qui concerne l'imagerie visuelle par exemple,

« si je ferme les yeux et je visualise, disons, le visage de ma femme, il semble entièrement faux de suggérer que le fait que ce que je visualise est le visage de ma femme – un fait qui met en relation mon état mental et l'environnement extra-psychologique – est étranger aux contenus de ma conscience, étranger à ce que je trouve quand j'"interromps le cours de ma pensée" »⁷¹.

Il faut distinguer trois versants de cette objection. Le premier versant de l'objection est épistémologique : faire des objets de l'introspection des objets mentaux serait incompatible avec une conception satisfaisante de la manière dont on connaît nos propres pensées. En effet, quand on pense, on sait ce qu'on pense immédiatement, c'est-à-dire non seulement sans observation du comportement, mais aussi sans processus d'interprétation. Quand on sait ce qu'on pense de cette manière, on le sait – semble-t-il – par introspection. Or, savoir ce qu'on pense, c'est savoir quels concepts on déploie et quel contenu on entretient. Donc, il paraît naturel d'affirmer que l'introspection nous livre des concepts et contenus, et pas seulement des objets mentaux ou présentations mentales. D'après cette objection, accepter la thèse de Putnam reviendrait à nier cet accès immédiat – non-interprétationnel – au contenu de nos pensées.

Le deuxième versant de l'objection est métaphysique, au sens où il concerne la nature des véhicules représentationnels, et la source de leur intentionnalité. Le cœur en est l'idée qu'un véhicule est par nature une entité dont le contenu dépend de l'interprétation. Ainsi, les représentations publiques, comme les énoncés linguistiques,

70 McDowell J., PMM, p. 44.

71 *Ibid.*, p. 44.

ont une intentionnalité dérivée⁷² – il faut y « lire » une signification, comme dit McDowell. Or l'intentionnalité de nos propres états mentaux n'est pas elle-même dérivée (et d'ailleurs, ne pourrait sans doute pas l'être sous peine de régression). Donc, avoir des états mentaux ne consiste pas à manipuler des véhicules. Si une telle objection vise avant tout la théorie représentationnelle de la pensée, elle s'applique aussi à la conception putnamienne des objets de l'introspection, si on interprète celle-ci comme identifiant les objets mentaux à des véhicules. En effet, d'après cette objection, il est également faux de considérer ces derniers comme nécessitant une interprétation.

Enfin, le troisième versant de l'objection est phénoménologique. La conception putnamienne des objets de l'introspection ne refléterait pas notre expérience vécue, qui est tout à fait différente d'une expérience de véhicules. Putnam soutient, on l'a vu, qu'il est concevable que des doubles introspectifs ne soient pas des doubles conceptuels. À cela, McDowell répond par ce qu'on a appelé par la suite⁷³ un argument du *contraste phénoménal* : l'effet que cela fait d'avoir en tête un certain mot n'est pas le même selon que l'on comprenne ou non ce mot en y associant le concept qu'il exprime. De même, l'expérience d'une image est différente quand on identifie conceptuellement ce dont c'est l'image. Ainsi il y a, affirme McDowell, une « phénoménologie de la pensée [*phenomenology of thinking*] »⁷⁴. Selon cette thèse, les concepts ne sont pas simplement individués par leurs rôles fonctionnels/inférentiels, mais aussi par leurs caractères phénoménaux caractéristiques. Il s'ensuit que l'on ne peut pas vraiment concevoir des doubles phénoménaux qui ne sont pas des doubles conceptuels, ni l'inverse, puisque les différences conceptuelles correspondent systématiquement à des différences phénoménologiques. McDowell suggère même que toute différence non seulement de concept, mais même de contenu référentiel induirait une différence phénoménale – c'est-à-dire souscrire à une forme d'*externalisme phénoménal*⁷⁵ selon laquelle le

72 L'intentionnalité dérivée est, par définition, extrinsèque. Cependant, d'après la plupart des conceptions des propriétés essentielles, l'intentionnalité dérivée pourrait – en principe – être essentielle. Je remercie un relecteur anonyme pour m'avoir invité à clarifier ce point.

73 Siegel S., *The Contents of Visual Experience*, Oxford, Oxford University Press, 2010.

74 McDowell tel que je le lis anticipe donc la position de partisans de la phénoménologie cognitive, notamment celle de Pitt D. (« The Phenomenology of Cognition, Or, What Is It Like to Think That *P*? », in *Philosophy and Phenomenological Research*, n°1, 2005, p.1-36) qui prête aux concepts des phénoménologies caractéristiques qui les individualisent. Certaines des réponses à McDowell que je propose au nom de Putnam, sont parallèles à certaines de celles de Levine J. (« On the Phenomenology of Thought », in T. Bayne & M. Montague (dir.), *Cognitive Phenomenology*, Oxford, Oxford University Press, 2011, p. 103-120) qui répond à Pitt.

75 Voir McCulloch G., « Phenomenological Externalism », in N. Smith (dir.), *Reading McDowell*, London, Routledge, 2002, p. 123-139. Je remercie Bruno Gnassounou de m'avoir signalé les travaux de cet auteur.

caractère phénoménal de nos pensées serait constitué en partie par leurs référents. L'externalisme phénoménal est incompatible avec la concevabilité non seulement de doubles phénoménaux qui ne soient pas des doubles conceptuels, mais même de doubles phénoménaux dont les pensées auraient des référents distincts.

6. Réponses aux objections de McDowell

S'il n'est pas possible ici de critiquer en détail l'externalisme radical de McDowell en général, ni son externalisme phénoménal en particulier, il me semble qu'il est néanmoins possible de montrer que les objections qu'on vient d'exposer reposent en partie sur une interprétation peu charitable de la position de Putnam, et qu'il est possible d'y apporter des réponses.

Commençons par répondre au versant épistémologique de l'objection. Celle-ci repose sur une ambiguïté. Que veut-on dire quand on dit que l'on sait immédiatement ce que l'on pense ? Distinguons trois sens de « savoir ce que l'on pense », qui correspondent à différentes notions de « conscience »⁷⁶ :

Je peux savoir que je pense que *p* au sens où je suis en mesure de former une pensée métareprésentationnelle de la forme : *Je pense que p*. Disons dans ce cas que ma pensée de premier ordre (que *p*) est *m-consciente*⁷⁷.

Je peux savoir que je pense que *p* au sens plus minimal où j'ai un *accès conscient*⁷⁸ à ce que je pense : ma pensée que *p* est librement disponible pour un grand nombre d'autres processus rationnels, tels que le contrôle de l'action, le raisonnement (je peux tirer des conclusions qui suivent de la prémisse que *p*), et typiquement la parole (je peux rapporter linguistiquement que « *p* »). On dira dans ce cas que ma pensée est *a-consciente*. Le statut *a-conscient* d'un état mental est une propriété relationnelle qui dépend de son rôle fonctionnel, c'est-à-dire de ses relations causales/inférentielles avec les autres états.

Enfin, un état mental peut être conscient au sens *phénoménal* quand il y a quelque chose que cela fait d'être dans cet état. On dira que l'état est *p-conscient*. Nous avons vu

76 La distinction bien connue entre conscience d'accès et conscience phénoménale provient de Block N., « On a Confusion about a Function of Consciousness », in *Brain and Behavioral Sciences*, vol. 18, n°2, 1995, p. 227-247.

77 Voir Smithies D., « A Simple Theory of Introspection », in D. Smithies & D. Stoljar (dir.), *Introspection and Consciousness*, Oxford, Oxford University Press, 2012, p. 259-294 et Block, *op.cit.* On pourrait étendre la notion de *m-conscience* pour inclure des formes procédurales de *métacognition* (voir Proust J., *The Philosophy of Metacognition: Mental agency and Self-Awareness*, Oxford, Oxford University Press, 2013). Je remercie J. Smortchkova pour avoir attiré mon attention sur ce point.

78 Block N., *op.cit.*

que les états p-conscients paradigmatiques sont des états perceptuels ou sensoriels, qui ont de manière relativement peu controversée des caractères phénoménaux caractéristiques. Mais on pourrait envisager que les pensées en un sens plus strict, c'est-à-dire les attitudes propositionnelles doxastiques, comme les jugements et croyances, soient aussi p-conscientes. Il me semble que c'est précisément la thèse de McDowell, quand il parle d'une « phénoménologie de la pensée » et attribue une phénoménologie aux concepts.

Armés de ces distinctions, nous pouvons expliquer en quoi l'objection de McDowell repose sur une ambiguïté. Pris en certains sens, affirmer que l'on sait immédiatement ce que l'on pense est une platitude : c'est une platitude d'affirmer que nous sommes en mesure de former des pensées a-conscientes et m-conscientes sans passer par l'observation du comportement ou par un processus (conscient) d'interprétation. Mais rien de ce que dit Putnam ne contredit cela. En particulier, rien de ce que dit Putnam ne suggère que, pour être a-conscient ou m-conscient de nos pensées, il faille en passer par l'interprétation de nos objets mentaux ou présentations mentales.

Ainsi, une pensée est a-consciente simplement en vertu du fait qu'elle joue le bon rôle fonctionnel dans notre cognition, ou pour parler comme Putnam dans RVH, qu'elle correspond à l'exercice de certaines capacités. Si on a une certaine pensée, et qu'on a les capacités inférentielles qui correspondent à son contenu sémantique (on peut en tirer des conséquences logiques évidentes, etc.), alors on sait immédiatement ce que l'on pense en ce sens. Ainsi, contrairement à ce que pourrait laisser supposer l'objection de McDowell, la position de Putnam n'est nullement incompatible avec l'idée que quand j'entends de l'eau qui coule, je peux en tirer immédiatement la conséquence que ce que j'entends est de l'eau. En effet, je peux faire cela simplement en ayant la pensée a-consciente que j'entends de l'eau.

De même, rien de ce que dit Putnam n'implique que l'on ne puisse pas savoir ce que l'on pense, au sens d'en former une métareprésentation sans en passer par l'interprétation d'un objet mental. Même si Putnam lui-même ne s'attarde pas sur ce point, voici par exemple une manière simple dont une pensée a-consciente pourrait devenir m-consciente de manière non-interprétationnelle dans un cadre théorique fonctionnaliste semblable au sien. Disons, de manière imagée, que la pensée occurrente que *p* figure dans ma « boîte de croyances »⁷⁹. Je suis alors automatiquement disposé, en vertu d'un mécanisme subpersonnel, à lui préfixer « Je crois que ... ». De sorte que j'en

79 Schiffer S., « Truth and the Theory of Content », in J. Bouveresse & H. Parret (dir.), *Meaning and Understanding*, Berlin, Walter de Gruyter, 1981, p. 204-222.

viens à croire de manière fiable que je crois que p quand je crois que p ⁸⁰. Ainsi, je peux savoir ce que je pense – au sens d'en être m-conscient – de manière non-interprétationnelle. Putnam pourrait ainsi répondre à McDowell que rien dans sa position ne contredit l'idée selon laquelle quand j'entends de l'eau qui coule, je suis généralement disposé à former immédiatement la croyance *que j'entends de l'eau qui coule*.

De même, une part de la force intuitive de l'objection McDowellienne à l'expérience de pensée de l'illusion introspective repose sur une ambiguïté possible dans la manière d'interpréter les états métareprésentationnels qu'elle met en jeu. Il faut soigneusement distinguer l'hypothèse de Putnam de la concevabilité des illusions de pensée de l'hypothèse voisine, moins plausible, selon laquelle je pourrais *croire que je crois que p* , mais pourtant *ne pas croire que p* . Cette hypothèse paraît peu plausible dans la mesure où la croyance *que je crois que p* semble contenir comme constituant la croyance de premier ordre *que p* , et donc impliquer que je possède effectivement les concepts qui composent p . Mais bien comprise, l'expérience de pensée de Putnam correspond plutôt à un scénario où j'ai une attitude envers *la phrase mentale « p »*, sans avoir la croyance correspondante *que p* . Et la thèse de Putnam est précisément que la phrase mentale n'est pas composée de concepts.

En quel sens alors est-ce que Putnam nie effectivement que mes concepts et les contenus de mes pensées soient introspectibles ? Il me semble qu'il s'agit uniquement du sens phénoménal. Quand Putnam dit qu'on introspecte un état, il veut dire non seulement que cet état est a-conscient et en mesure d'être m-conscient, mais en outre qu'il y a un effet caractéristique que cela fait d'être dans cet état – c'est-à-dire que c'est un état constitutivement p-conscient. Dit autrement, Putnam soutient que des sujets ne peuvent être des doubles introspectifs que s'ils sont des doubles phénoménaux. Ainsi, ce serait (selon mon interprétation⁸¹) une condition nécessaire pour qu'une attitude propositionnelle soit introspectible au sens pertinent *pour Putnam*, que cette attitude ait un caractère phénoménal caractéristique⁸². À mon avis, ce qu'il nie, en niant que les

80 Voir Levine, *op.cit.*

81 Un argument en faveur de cette interprétation est qu'elle est tout à fait en accord avec le contexte plus large dans lequel s'inscrit la discussion des objets de l'introspection. Comme on l'a vu, il s'agit pour Putnam de s'opposer aux théories phénoménologiques ou empiristes qui confondent concepts et objets mentaux, c'est-à-dire qui individuent les concepts par des propriétés sensorielles ou phénoménales.

82 L'« introspection » au sens strict où l'entend Putnam, ne désigne donc pas n'importe quelle méthode de connaissance de soi permettant de savoir ce qu'on pense – en être a-conscient ou m-conscient – sans passer par l'observation du comportement ou l'interprétation, mais plutôt le type d'accès immédiat qu'on a à nos états phénoménaux. Ainsi, Putnam ne nierait pas qu'on puisse connaître le contenu de nos états

concepts ou contenus eux-mêmes soient introspectibles, c'est donc simplement qu'on en soit conscient comme, ou en vertu du fait, qu'on soit conscient de certaines propriétés phénoménales.

À cela, McDowell aurait pu répondre en niant que la p-conscience soit nécessaire à l'introspectibilité. Mais, comme on l'a vu, ce n'est pas le cas. À la place, McDowell défend la thèse de la phénoménologie de la pensée. Il accepte ainsi apparemment que seuls les états p-conscient sont introspectibles, mais il insiste sur le fait que les contenus intentionnels et les concepts *sont* caractéristiquement p-conscients, en sus des états sensoriels ou perceptuels. À ce stade, le versant épistémologique de l'objection de McDowell rencontre donc son versant phénoménologique : la question devient celle de savoir si l'on sait immédiatement ce que l'on pense au sens précis où on le sait *sur la base de son caractère phénoménal propre*. Mais, comme on le verra par la suite, il n'est pas du tout clair que la pensée conceptuelle ait en tant que telle un caractère phénoménal. Ce qui suffit, pour l'heure, à répondre au versant épistémologique de l'objection. C'est seulement en un sens phénoménologique très particulier que Putnam nie que l'on puisse introspecter les concepts et contenus.

Ainsi, le versant épistémologique de l'objection de McDowell tire sa force apparente de l'ambiguïté entre la platitude que l'on est immédiatement a-conscient ou m-conscient de ses pensées, et la thèse controversée selon laquelle on en est p-conscient. Que la théorie de Putnam implique le rejet de cette dernière thèse ne la rend nullement incompatible avec une théorie sinon évidemment correcte, du moins éminemment plausible de la manière dont on connaît ses propres états mentaux, contrairement à ce que suggère McDowell.

Répondons à présent au deuxième versant de l'objection de McDowell : est-il vrai que les représentations mentales, en tant que véhicules, soient le type d'entités qu'il faille nécessairement interpréter pour comprendre, de sorte qu'elles ne puissent recevoir un contenu que de manière dérivée ? C'est une pétition de principe, puisqu'une telle thèse est généralement rejetée par les défenseurs de la théorie représentationnelle de la pensée. Selon eux, il s'agit là d'une différence cruciale entre les représentations publiques (par exemple linguistiques) et les représentations mentales : il est faux que « comprendre » ce que l'on pense revienne à *interpréter* ses véhicules mentaux, ou que le

mentaux en suivant la méthode dite de « transparence » (Evans G., *op.cit.*). Selon cette méthode, je peux répondre à la question de second-ordre : « Est-ce que je crois que *p* ? » simplement en répondant à la question de premier ordre : « Est-ce que *p* ? ». Même si une telle méthode ne passe pas par l'interprétation ou l'observation du comportement, elle n'est clairement *pas* introspective au sens de Putnam.

contenu des représentations *mentales* soit nécessairement dérivé⁸³. En effet, on « comprend » les représentations mentales directement en pensant *avec*, c'est-à-dire dans la mesure où elles jouent le bon rôle fonctionnel/inférentiel, plutôt qu'en pensant à ces représentations, c'est-à-dire en formant des métareprésentations leur attribuant un contenu.

Mais même si une telle réponse s'applique aux véhicules de pensée, la position de Putnam ne fait-elle pas face à l'objection selon laquelle les *objets mentaux* non plus n'ont pas besoin d'être interprétés ? Une première réponse est, comme on l'a vu, que les objets mentaux ne sont tout simplement pas des véhicules, mais des représentations de véhicules (publics) dans le contenu de nos pensées. On est immédiatement a-conscient qu'un visage imaginé est celui de notre femme précisément de la même manière qu'on est immédiatement a-conscient de ce qui est représenté dans le contenu de ce qu'on imagine en général.

L'autre réponse possible correspond à la deuxième conception possible de la parole intérieure précédemment mentionnée, à savoir celle selon laquelle parler intérieurement, c'est parler sans vocaliser. Quand je parle, j'ai un accès non-interprétationnel au sens de ce que je dis parce que c'est moi qui parle, c'est-à-dire que c'est moi-même qui produis intentionnellement mes paroles, et que je sais ce que j'ai l'intention de dire – au sens où j'en suis a-conscient. Il ne s'ensuit nullement que mes paroles ne soient pas des paroles, mais plutôt des entités essentiellement sémantiques (comme des sens frégréens). Pourquoi la même chose ne vaudrait-elle pas *mutatis mutandis* pour les paroles intérieures ?

Venons-en au troisième versant de l'objection, c'est-à-dire son versant phénoménologique. Notons pour commencer qu'il est faux que la thèse identifiant ontologiquement les objets de l'introspection à des objets mentaux implique que ceux-ci doivent nous apparaître toujours phénoménologiquement *en tant que véhicules*. Quand on voit un trompe-l'œil, on ne voit guère que c'est une image que l'on voit. Ainsi, on ne voit pas (au sens référentiellement opaque de « voir ») les propriétés formelles de l'image, mais ses propriétés intentionnelles, c'est-à-dire ce dont c'est une image⁸⁴. Au sens référentiellement transparent, l'objet vu reste une image. On pourrait en principe soutenir une thèse analogue au sujet de l'introspection. Ainsi, même la thèse – à mes

83 Fodor J., *The Language of Thought*, *op.cit.*

84 La perception des images physiques donne lieu à une expérience « double » [*twofold*] caractéristique, où l'on perçoit l'image/la surface en tant que telle *et/ou* son contenu – ainsi, on voit *dans* l'image (Wollheim R., « On Pictorial Representation », in *Journal of Aesthetics and Art Criticism*, vol. 56, n°3, 1998, p. 217-226).

yeux peu plausible – selon laquelle il y aurait une *transparence phénoménale* des objets de l'introspection (par analogie avec la transparence phénoménale de la perception⁸⁵), c'est-à-dire selon laquelle il serait impossible d'accéder à quelques propriétés non-intentionnelles des objets de l'introspection que ce soit, n'impliquerait pas que ce que l'on introspecte (au sens référentiellement transparent) ne soient pas en fait des véhicules mentaux.

Quoi qu'il en soit, le cœur du versant phénoménologique de l'objection de McDowell ne consiste pas à soutenir que les objets de l'introspection sont phénoménalement transparents. Il s'agit plutôt de soulever la question cruciale du rapport entre les propriétés phénoménales de nos pensées, et leurs contenus intentionnels ou conceptuels. Selon Putnam, les pensées et les concepts sont accompagnés d'images verbales ou mentales. La phénoménologie de la pensée est en fait la phénoménologie des états perceptuels, sensoriels, ou émotionnels *associés* à la pensée, et non la phénoménologie *de la pensée* elle-même. Ainsi, Putnam ne nierait nullement que quand je pense à de l'eau, cela me fait un certain effet expérientiel, au sens où j'ai en pensant une riche expérience subjective. Mais selon lui, exercer le concept eau ou faire référence à l'eau ne *consiste* pas à avoir une expérience phénoménale d'un certain type. En particulier, les conditions d'individuation des concepts et des contenus ne sont pas phénoménologiques, ce qui explique que des doubles phénoménaux puissent ne pas être des doubles conceptuels, et vice-versa. Selon McDowell, au contraire, la phénoménologie de nos épisodes de pensée ne se réduit pas à la phénoménologie sensorielle, affective, etc. associée à la pensée, mais est une phénoménologie propre à la pensée elle-même. Exercer un concept ou faire référence, c'est – entre autres choses – nécessairement avoir une certaine expérience phénoménale. Il s'ensuit que les expériences de pensée de Putnam où pensée et phénoménologie sont dissociées ne sont pas réellement concevables, du point de vue McDowellien.

Je ne prétends pas ici trancher entièrement le débat de savoir s'il y a ou non une phénoménologie de la pensée au sens fort de McDowell, ou seulement au sens faible de Putnam. Je me contenterai de noter une faiblesse majeure de l'argument de McDowell. Il ne suffit pas de relever que l'effet que cela me fait de visualiser un visage ou un mot quelconque est tout à fait différent de l'effet que cela me fait de visualiser un visage que je conceptualise comme celui de ma femme, ou un mot que je comprends, pour pouvoir en conclure que le concept ma femme ou la signification du mot contribue *directement*

85 Harman G., « The Intrinsic Quality of Experience », in *Philosophical Perspectives*, vol. 4, 1990, p. 31-52.

au caractère phénoménal de l'état dans lequel je me trouve. En effet, il faut montrer que le contraste phénoménal global entre les états contrastés – ceux avec et sans déploiement conceptuel – ne s'explique pas tout aussi bien par la contribution phénoménale des états sensoriels, perceptuels, etc. *associés* aux concepts. Ainsi, quand j'exerce le concept MA FEMME, s'active un riche réseau de représentations – les images mentales, mots mentaux, sentiments, etc. qui constituent ma conception (et non le concept) de ma femme ou dérivent de celle-ci. Par exemple, j'aurais typiquement à l'esprit une représentation du nom de ma femme, en même temps que son visage – c'est-à-dire un certain mot mental p-conscient. Une telle explication du contraste phénoménal paraît raisonnable, et est compatible avec l'hypothèse selon laquelle seuls les états qualitatifs sensoriels ou perceptuels ont des caractères phénoménaux propres. Or, McDowell ne donne tout simplement pas d'argument contre cette hypothèse, et ne fait rien pour montrer que la phénoménologie qu'il considère comme caractéristique de nos concepts ou du contenu intentionnel de nos pensées n'est pas simplement la phénoménologie des objets mentaux ou présentations mentales qui les accompagnent. En l'absence d'un tel argument, son objection phénoménologique échoue.

Avant de conclure, il nous faut examiner une dernière objection possible contre Putnam. Celle-ci est une objection à l'argument de l'illusion introspective qui n'est pas formulée explicitement par McDowell, mais qui aurait pu l'être – au sens où une telle réponse me semble extrapolable à partir de ce qu'il dit contre l'interprétation en termes de la conception duplex de la pensée de l'expérience de pensée de Terre Jumelle⁸⁶. En effet, il y a – en apparence du moins – une parenté de structure entre l'argument de Terre Jumelle, l'argument classique de l'illusion en philosophie de la perception, et l'argument du sujet qui a l'illusion de comprendre son monologue intérieur. En tenant compte de ce que McDowell dit à propos des deux premières, on peut extrapoler ce qu'il aurait pu dire à propos de la troisième.

Dans l'argument classique de l'illusion, deux sujets ont des expériences visuelles subjectivement indistinguables, mais seule l'expérience de l'un des deux s'explique par les objets mondains qui lui sont présentés visuellement, puisque l'autre est sujet à une illusion/hallucination. On en conclut que le caractère phénoménal de l'expérience ne s'explique pas par les objets mondains effectivement perçus, mais par un autre facteur, commun au « bon cas » (la perception réussie) et au « mauvais cas » (l'illusion/hallucination). McDowell fournit une réponse *disjonctiviste*⁸⁷ à cet argument :

86 McDowell J., PMM, p. 41.

87 McDowell J., « Singular Thought and the Extent of "Inner Space" », *op. cit.*

il rejette le présupposé que le bon cas et le mauvais cas sont fondamentalement de même nature. McDowell étend une telle réponse à l'argument de Terre Jumelle, si on l'interprète comme un argument pour la conception duplex de la pensée. Un tel argument présuppose de manière analogue qu'il faille concevoir le bon cas (la pensée sur Terre) et le mauvais cas (la pensée sur Terre Jumelle) comme fondamentalement de même nature, car résultant d'un facteur commun aux deux cas – à savoir le facteur étroit postulé par la conception duplex.

On pourrait être tenté d'identifier un présupposé semblable dans l'argument de l'illusion introspective de Putnam. Encore une fois, nous avons deux sujets similaires selon une certaine dimension subjective – en l'occurrence, des doubles introspectifs – mais qui diffèrent selon une autre dimension – en l'occurrence, la dimension conceptuelle. Le sujet dans le mauvais cas a l'impression d'avoir accès à des concepts, mais n'a en fait accès qu'à des objets mentaux. Putnam a l'air d'en tirer la conclusion qu'il faille expliquer l'expérience introspective des sujets dans les deux cas par un facteur commun – en l'occurrence, les objets mentaux.

Dans une veine McDowellienne⁸⁸, on pourrait répondre que l'argument de l'illusion introspective ainsi formulé souffre du présupposé que les deux cas sont fondamentalement de même nature, et qu'il généralise ainsi abusivement à partir de ce qui se passe dans le mauvais cas. Certes, dans le cas où a lieu l'illusion, le sujet a un accès introspectif à un objet mental qu'il prend à tort pour l'exercice d'un concept. Mais cela n'implique pas que dans le bon cas, le sujet n'ait pas accès directement à ses concepts, plutôt qu'à ses objets mentaux.

En réponse, je me contenterais de noter que le parallèle entre l'argument de l'illusion perceptuelle et l'argument de l'illusion introspective est en fait bien plus limité qu'une telle objection pourrait laisser croire.

D'abord, alors que l'existence même du type d'entités censées constituer le facteur commun entre illusion et perception est controversée (par exemple, les *sense data*), l'existence de ce que Putnam appelle les « objets mentaux » est de bon sens, et n'est nullement niée par McDowell, qui parle lui-même d'imagerie mentale.

Ensuite, il convient de songer à l'autre expérience de pensée de Putnam – celle du penseur sans monologue intérieur. Comme le montre celle-ci, Putnam ne soutient aucunement que notre accès à nos propres pensées soit toujours médiatisé par des objets mentaux, ni même que nos pensées s'accompagnent nécessairement de telles

88 Une position disjonctiviste au sujet de l'introspection semblable à celle que l'on esquisse ici est développée par Macpherson, *op. cit.*, qui l'applique au syndrome d'Anton.

entités. Au contraire, selon lui, les objets mentaux ne jouent tout simplement aucun rôle *constitutif* dans la pensée, et, comme on l'a vu, ne sont pas à identifier avec le premier facteur dans une conception duplex de la pensée. Putnam pourrait donc maintenir sa thèse selon laquelle les objets de l'introspection – au sens strict où celle-ci nécessite la p-conscience – se limitent à des objets mentaux, tout en acceptant l'idée qu'être dans un certain état conceptuel et avoir faussement l'impression de l'être sont des états de natures fondamentalement distinctes. Tel que je l'interprète, donc, l'argument de l'illusion introspective n'a pas pour but d'introduire un « voile » d'objets mentaux entre le sujet et ses concepts, mais plutôt d'établir une distinction nette entre concepts et objets mentaux, tels que les images et les mots qui traversent notre flux de p-conscience.

7. Conclusion

J'espère avoir montré que la thèse de Putnam dans RVH selon laquelle les objets de l'introspection sont des objets mentaux ou des présentations mentales plutôt que des concepts demeure plausible – ou du moins qu'elle résiste aux objections de McDowell dans PMM. J'espère également avoir montré que cette thèse est à distinguer soigneusement d'autres thèses endossées par Putnam, c'est-à-dire la théorie représentationnelle et la conception duplex de la pensée, dont elle est indépendante.

Je ne prétends nullement avoir établi ces dernières, ni bien sûr avoir réfuté l'externalisme radical de McDowell. Cependant, j'aimerais en guise de conclusion revenir rapidement sur les raisons, que donne assez allusivement celui-ci dans PMM, pour lesquelles la théorie représentationnelle devrait être abandonnée. Comme on l'a vu, McDowell suggère qu'abandonner la théorie représentationnelle aurait l'avantage de dissoudre le problème de la référence. Cela n'est pas le cas. Le problème de la référence présuppose qu'il conviendrait d'expliquer de manière substantielle la nature de la référence, en termes d'une autre relation. Mais ce problème ne présuppose rien en ce qui concerne la nature exacte des entités qui réfèrent. En particulier, il ne présuppose nullement que les premiers termes de la relation de référence soient des véhicules représentationnels internes aux individus, semblables aux symboles. Ainsi, supposons, en tirant inspiration de McDowell, que seules les personnes toutes entières, inscrites dans un contexte social et imbriquées dans un riche tissu de relations normatives, causales, etc. avec leur environnement, soient susceptibles de référer – c'est-à-dire d'accomplir des *actes* de référence. On pourrait néanmoins persister à demander *en vertu de quoi* de tels actes de référence sont possibles, et insister pour qu'il ne faille pas

que la réponse fasse tacitement appel à la notion de référence elle-même, sous peine de circularité. En d'autres termes, le problème de la référence ne présuppose pas la théorie représentationnelle de la pensée, mais seulement que la relation de référence *n'est pas fondamentale* – ce qui est une thèse certes contestable, mais tout à fait indépendante⁸⁹.

Ce qui nous amène à l'autre raison que donne McDowell de rejeter la théorie représentationnelle : d'après McDowell, il s'agit d'un reliquat de « scientisme⁹⁰ ». Cette idéologie nous pousserait à endosser la conception de la pensée qui semble en faire un objet d'étude scientifique pour la psychologie ou les sciences cognitives, en assimilant la pensée à l'activité d'un « organe ». Il me semble pour ma part que le fait que la théorie représentationnelle de la pensée soit présupposée par nos meilleures théories empiriques de l'esprit est, quoi qu'en dise McDowell, un sérieux argument en faveur de cette théorie. Mais je tiens à souligner l'existence d'au moins un argument sérieux pour les véhicules de pensée qui s'appuie sur une prémisse *frégéenne*, plutôt que sur un quelconque scientisme. Suivant Frege, il nous faut distinguer les référents de nos pensées de leurs modes de présentation. Ceux-ci obéissent à la contrainte de Frege⁹¹, selon laquelle si un sujet peut sans contradiction interne attribuer au même référent des propriétés contradictoires, c'est qu'il attribue ces propriétés à l'objet sous des modes de présentations distincts. Or une différence de mode de présentation semble compatible avec une identité de *contenu descriptif*⁹². Ceci suggère que les différences dans les valeurs cognitives des pensées doivent parfois s'expliquer en termes de *modes de présentation non-descriptifs*, que certains philosophes proposent d'identifier à des *véhicules mentaux*⁹³. Ici n'est pas le lieu pour discuter de cette position en détail, mais il paraît faux de dire comme McDowell que seul un « scientisme résiduel » aurait pu pousser Putnam à accepter la théorie représentationnelle. Cette théorie, que Putnam contribua remarquablement à développer, reste indépendamment prometteuse. Ainsi, non seulement la conception des objets de l'introspection de Putnam à l'époque de RVH, mais

89 S. Echeverri développe de manière convaincante un thème et un argument assez semblables dans « Explaining Reference: A Plea for Semantic Psychologism », in J. Dutant, D. Fassio & A. Meylan (dir.), *Liber Amicorum Pascal Engel*, University of Geneva, 2014, p. 550-580.

90 McDowell J., PMM. p. 46.

91 Schiffer S., « The Mode-of-Presentation Problem », in C. A. Anderson & J. Owens (dir.), *Propositional Attitudes: The Role of Content in Logic, Language, and Mind*, Stanford, CSLI, 1990, p. 249-268.

92 Voir notamment Fine K., *Semantic Relationism*, London, Blackwell, 2007.

93 Voir par exemple la théorie des fichiers mentaux (Recanati F., *Mental Files*, Oxford, Oxford University Press, 2012 ; Fodor J., *LOT 2: The Language of Thought Revisited*, Oxford, Oxford University Press, 2008). Cette théorie, en identifiant les modes de présentation non-descriptifs à des véhicules, s'oppose à la conception néo-frégéenne de ceux-ci proposée par Evans ou McDowell.

encore sa vision plus globale de la nature de la pensée, demeurent entièrement d'actualité.