

ADHÉSION ET JUSTIFICATION

[Jean-Baptiste Rauzy](#)

Article paru dans « [Les Études philosophiques](#) »

2018/4 N° 184 | pages 585 à 609

Version auteur. La pagination de l'éditeur a été intégrée.

ADHÉSION ET JUSTIFICATION

1. De la conviction à l'adhésion

Dans les pages qu'il a consacrées à la phénoménologie de la conviction, Fernando Gil a revendiqué la filiation avec Fichte. Il a trouvé chez Fichte une description de l'esprit comme tendance (*Trieb*) et agilité ; du mouvement qui s'accomplit et s'achève dans l'acte de connaissance et dans la saisie d'un objet. Tendance, mouvement et agilité sont également présents dans l'expérience de la croyance et dans celle de la conviction ; mais, écrivait Gil, dans la seconde, la conviction, l'acte de connaissance se superpose à la tendance et vient l'achever, alors que la simple croyance institue son objet, confond, pour ainsi dire, l'opération de l'esprit et l'objet qu'elle engendre¹. Reprenant, peut-être plus lointainement, la tradition du *constituit intellectum*², Gil propose de considérer que :

le sujet S croit que *p*

signifie principalement qu'il est *porté* à le croire. Alors que :

le sujet S est convaincu que *p*

1. « La conviction porte sur une entité, une relation, un prédicat, qui se dédouble en une contrainte d'acceptation. Subjectivement, la conviction est cette contrainte même, l'adhésion investissant son objet. La croyance a, elle aussi, une teneur objective et mobilise la même adhésion, mais l'investissement anticipe son objet en instituant le réel auquel on croit [...] La croyance s'épargne la détermination de l'objet par les opérations de connaissance. De la tendance on saute directement à l'acte, la croyance confondant l'opération et l'objet qu'elle engendre » (Fernando Gil, *La Conviction*, Paris, Flammarion, 2000, p. 127).

2. Aristote précise par parenthèse qu'il y a production de signification lorsque celui qui parle peut arrêter son intelligence. La phrase grecque est ainsi rendue dans le latin de Boèce : « Ipsa quidem secundum se dicta verba nomina sunt et significant aliquid – constituit enim qui dicit intellectum, et qui audit quiescit » (*Aristoteles Latinus*, II, 1-2, p. 7). Dans le commentaire de ces lignes 16b19-25 du *De Interpretatione*, Boèce a pris très au sérieux la métaphore dynamique du mouvement et du repos et il entendait son propre « *constituere* » comme un « fixer » ou « arrêter ». Cf. Thomas d'Aquin, *In Peri Hermeneias*, L I, l. V, 68. L'intellection est un mouvement de l'âme qui se cherche un repos et qui trouve un concept.

signifie que S est porté à croire que p mais possède en outre de bonnes raisons tirées de la constitution d'un objet. La distinction entre croyance et conviction repose, selon Gil, sur la contribution de l'objet à l'attitude doxastique du sujet.

Une objection qui vient aussitôt à l'esprit porte sur le rôle respectif des conditions de vérités et des conditions de justification dans cette distinction. Il n'est pas toujours aisé de savoir quelles sont les limites de l'objet qui contribue à la conviction lorsque celle-ci est donnée en termes propositionnels. Supposons que le contenu de l'attitude soit, par exemple, une propriété morale :

Anna est convaincue que l'être humain est cupide

Supposons également qu'Anna soit assez informée, par exemple une sociologue chevronnée qui a tiré de son travail de bonnes raisons de penser que l'être humain, placé dans des conditions de vie concrètes, ne peut pas, pour des raisons structurelles, ne pas développer une forme de cupidité. Supposons en outre que ce soit effectivement le cas ; autrement dit que l'être humain soit effectivement cupide et donc que la croyance/conviction d'Anna soit vraie. Il se peut néanmoins qu'Anna, malgré ses années d'études et de recherches, adhère à ce contenu pour une raison plus personnelle, qui tient à un détail de sa biographie, au caractère de son père ou à ses expériences et déconvenues, par exemple, lors de son divorce. Dans ce cas l'objet, si l'on peut dire, qui contribue à la conviction d'Anna, n'est pas exactement celui qui intervient dans les conditions de vérité du contenu envisagé. Nous pouvons espérer nous mettre d'accord sur ce qui doit être le cas pour que « l'être humain est cupide » soit vrai en dépit du caractère vague de la notion de cupidité. En revanche, la question de savoir pour quelle raison tel sujet particulier en est convaincu est plus obscure. Le sujet lui-même – ou son biographe – pourrait modifier ses vues sur ce point et parvenir à une conclusion montrant que les raisons contingentes de croire sont très différentes des raisons de tenir pour vrai. Les conditions de vérité d'un contenu mental propositionnel ont un caractère public partagé. Les conditions de la conviction sont plus variables et insaisissables. La croyance est sur ce point assez équivoque et regarde pour ainsi dire des deux côtés, au point que certains épistémologues ont souhaité s'en débarrasser. Keith Lehrer, par exemple, a proposé très tôt de lui préférer une forme de *acceptance*³ dans la définition de la connaissance, c'est-à-dire la croyance, l'attitude doxastique, mais en tant qu'elle est essentiellement tournée vers la vérité.

L'option de Gil paraît d'autant plus incertaine et paradoxale. Elle consiste à creuser le côté subjectif de la croyance dans l'espoir d'y trouver une source

3. K. Lehrer, *Theory of Knowledge*, Boulder, Westview Press, 1990, pp. 35-36. La distinction entre *belief* et *acceptance* telle qu'elle a été introduite par Jonathan Cohen n'est pas tout à fait la même (voir L. J. Cohen, « Belief and Acceptance », *Mind*, 98/391, 1989, pp. 367-389).

stable du savoir objectif. Pourquoi s'intéresser à la conviction ? Pourquoi ne pas s'en tenir à la vérité ? Fernando Gil répond : la nature de l'aspect subjectif de la croyance, ses motifs, son lien à la justification ne sont pas indifférents à la constitution de l'objet et de l'objectivité. Je formule à dessein cette thèse de manière très générale, non pas pour suggérer que Gil aurait procédé de la même manière, mais plutôt pour souligner que l'héritage de Gil est lui-même assez large et peut être évoqué sous cette forme d'interrogation. La preuve, dans sa version mathématique ordinaire ou formellement élaborée, produit-elle de la conviction ? N'est-ce pas au contraire l'adhésion de l'esprit qui arrête la preuve ? Il m'a semblé qu'une manière de poursuivre l'enquête sur ces questions sans répéter ou commenter le parcours effectué par Gil lui-même est de la faire porter directement sur ce que j'ai choisi d'appeler « l'adhésion » et de chercher à cerner sa contribution épistémique.

2. Adhésion ou acquiescement ?

On ne rencontrera pas habituellement *l'adhésion* telle que je l'entends dans la littérature épistémologique ; c'est plutôt l'acquiescement à un contenu – *assent* – qui est discuté par les auteurs⁴ et constitue un composant de la croyance proche de celui que je cherche à mettre en relief – auquel il faut ajouter le sentiment (épistémique) qui, dans une tradition qui remonte à David Hume, est supposé accompagner l'acquiescement et la présentation d'un contenu auquel on a acquiescé⁵. Pourquoi s'écarter de cette habitude ? L'acquiescement est en général conçu sur le mode d'une affirmation ou d'un jugement. Il est dépendant d'une réponse. Il est même *identifié* à la réponse affirmative qui serait donnée par le sujet à une question, le plus souvent imaginaire, fabriquée sur la base du contenu acquiescé⁶. L'acquiescement de l'agent *S* à *p* est la réponse « oui » qui serait donnée par *S* à la question « est-ce que *p* ? » et, si *S* a donné cette réponse « oui », on suppose qu'une présentation de *p* s'accompagne pour *S* d'un sentiment correspondant. On suppose aussi, par un glissement insensible, que le même sentiment – ou un sentiment comparable – accompagne une présentation de *p* dans les cas où la réponse « oui » à la question « est-ce que *p* ? » n'est ni passée, ni présente mais seulement hypothétique ou possible. Enfin, dans le cas où l'on admet que les croyances ont une nature dispositionnelle, l'acquiescement est particulièrement requis pour les croyances occurrentes, il est même ce qui permet

4. Voir par exemple S. Shoemaker, « Moore's Paradox and Self-knowledge », in *The First-Person Perspectives and Other Essays*, Cambridge, Cambridge University Press, 1996.

5. Voir H. H. Price, *Belief*, Londres, George Allen & Unwin, 1969, avec les citations commentées de Hume aux pp. 165-169.

6. Voir par exemple K. P. Moser, *Knowledge and Evidence*, Cambridge, Cambridge University Press, 1989, p. 15 : « My believing that there is a fawn in the lawn requires my having assented, perhaps unconsciously, to the proposition that there is a fawn in the lawn. [...] One's genuinely assenting to a proposition is simply one's sincerely and understandingly affirming it. [...] Assenting requires subjective, psychological affirmation of a proposition. »

de les distinguer : un sujet est porté à croire que p – croyance dispositionnelle – si, lorsque certaines circonstances sont réunies, p fait l'objet de son acquiescement – croyance occurrente.

Il est aisé de constater que l'identification pure et simple de l'adhésion et de l'acquiescement aurait des conséquences malheureuses. Notons en premier lieu que, si l'adhésion à un contenu doit caractériser la croyance, elle peut également s'appliquer à sa version dispositionnelle. L'acquiescement de son côté ne permet pas de distinguer des croyances dispositionnelles en dehors de leurs conditions de manifestation. Admettons que je sois porté à croire que p . Il existe alors un ensemble de circonstances C dans lesquelles ma croyance en p se manifeste, mais il existe aussi au moins un contenu, disons q , que je *ne suis pas* porté à croire, quelles que soient les relations de C à l'actualité, c'est-à-dire également lorsque les circonstances C ne sont pas réunies. Par suite, lorsque ces circonstances C ne sont pas réunies, si néanmoins ma relation doxastique au contenu p doit être isolée et distinguée de celle que j'ai relativement à q – que je ne suis pas porté à croire – on voit mal comment faire cette différence des deux relations en termes d'acquiescement. Pourtant je suis porté à croire que p et non que q , à tenir p pour vrai, plus que q ; p fait partie de mes croyances, à la différence de q : l'adhésion permet d'aborder ces différences sans passer par la forme contestable d'une question ou par la dimension de l'assertion et du jugement. À moins qu'on préfère parler d'un *assent* implicite qui distinguera p de q en dehors des circonstances de leurs manifestations, mais c'est alors la fonction de l'acquiescement dans la démarcation des croyances dispositionnelles et des croyances occurrentes qui se trouve par là fragilisée. On notera en second lieu, que les conditions qui pèsent sur une élaboration de l'acquiescement ne s'imposent pas dans le cas de l'adhésion. Je pense d'abord à la *condition d'intelligibilité*. Il est raisonnable de prétendre qu'on ne peut pas authentiquement acquiescer à ce qu'on ne comprend pas alors qu'on peut y adhérer. Un étudiant ignorant peut adhérer au contenu [que la prose de Cicéron est pleine de synecdoques] sans savoir exactement ce qu'est une synecdoque et sans avoir lu plus de quelques pages du latin de Cicéron, si le professeur qui jouit d'un certain prestige à ses yeux le lui a assuré. Pour qu'il *acquiesce* en revanche, d'un acquiescement authentique, il faut au moins qu'il ait été capable de reconnaître une synecdoque, qu'il soit un tant soit peu versé dans la rhétorique et ait lu assez de Cicéron pour que la réponse qu'il pourrait donner à la question : « la prose de Cicéron est-elle pleine de synecdoques ? » soit pourvue de fondement. De même pour la *condition de sincérité*. Si notre étudiant répond « oui » à la question « la prose de Cicéron est-elle pleine de synecdoques ? » alors qu'il a lu Cicéron, sait ce qu'est une synecdoque mais, par extraordinaire, n'en a jamais rencontré au cours de ses lectures, dans ce cas on dira qu'il n'a pas véritablement acquiescé au contenu [que la prose de Cicéron est pleine de synecdoques], qu'il a plutôt dit « oui » pour se sortir d'affaire, mais il est toujours possible qu'il y adhère, parce que ce qu'il sait de Cicéron – qu'il

est un grand prosateur – et ce qu’il sait des prosateurs – qu’ils manient volontiers les figures du discours – le porte à croire qu’il en est ainsi.

Il y a par suite un ensemble de raisons solides qui nous invitent à distinguer ce qu’on peut exprimer au moyen de la métaphore de l’adhésion et l’*assent* proprement dit. Mais que répondra-t-on à celui ou celle qui remarque alors que si, comme on vient de le proposer, on met de côté l’*assent* ou acquiescement à un contenu, il ne reste rien dans la croyance qui puisse être visé par la métaphore de l’adhésion, si ce n’est éventuellement un – ou plusieurs – sentiments qui l’accompagnent ? De deux choses l’une, ou bien ce sentiment joue un rôle épistémique authentique et, dans ce cas, le fait qu’il s’agisse d’un sentiment est évidemment significatif, mais cela n’écarte pas l’adhésion-sentiment de l’agenda de l’épistémologue ; ou bien le sentiment en question est un produit secondaire peu important et, dans ce cas, on doit pouvoir maintenir qu’il y a encore besoin d’une composante supplémentaire et que c’est précisément celle que l’on entend chercher en termes d’adhésion. Ajoutons également que prendre en compte les sentiments épistémiques n’implique pas nécessairement de les relier à l’acquiescement. Car on peut accorder au humien que la représentation d’un contenu qui a fait l’objet d’un acquiescement n’est pas ressentie de la même manière, qu’elle est plus vive ou a plus de force, mais les sentiments qui indiquent qu’on peut, par exemple, compter sur un contenu, qu’on aura gardé des éléments pertinents en mémoire ou qu’on peut en faire une prémisse dans un processus inférentiel, tous ces sentiments sont plus larges et ne se limitent pas aux seuls contenus acquiescés. Lorsque je crois savoir que je vais retrouver le nom de telle personne croisée dans un colloque que j’ai connue il y a longtemps, je suis porté à chercher son nom parce que quelque chose – un sentiment – m’indique que je vais le retrouver et que mon effort n’aura pas été vain, mais il est peu probable que, dans ce cas, un contenu de croyance portant sur les chances que j’ai de retrouver le nom en question fasse *ipso facto* l’objet d’un acquiescement véritable⁷. Autrement dit : même si l’on accorde qu’il existe un sentiment qui trouve son origine dans l’acquiescement, cela ne signifie pas que tous les sentiments épistémiques qui favorisent nos attentes relatives à un contenu partagent cette origine.

3. Adhésion-signal et adhésion-couleur

Un mot à présent sur la phénoménologie de l’adhésion. Un même contenu doxastique peut être présenté de différentes manières, soit par des personnes différentes qui n’ont pas exactement les mêmes croyances, soit

7. La thèse selon laquelle les sentiments métacognitifs font interface entre des processus conscients et inconscients est défendue dans A. Koriat, « The Feeling of Knowing: some Metatheoretical Implications for Consciousness and Control », *Consciousness and Cognition*, 9, 2000, pp. 149-171 ; voir aussi J. Proust, « The Representational Structure of Feelings », in T. Metzinger et J. M. Windt (dir.), *Open Mind* : 31 (T), Francfort, MIND Group, 2015.

par un agent unique dont les croyances sont révisées dans le temps. Celle qui est convaincue que l'être humain est profondément cupide n'a pas la même relation au contenu propositionnel [que l'être humain est cupide] que celui, plus optimiste, qui ne le croit pas ou pense que l'être humain est cupide parfois mais pas toujours. Si l'adhésion affecte la représentation, alors la première se représente le contenu [que l'être humain est cupide] selon un mode de représentation caractéristique des contenus auxquels elle adhère, alors que la représentation du second pour le même contenu est colorée différemment. Il n'est pas indispensable d'être spécialiste de Hume pour avoir fait l'expérience de contenus que nous avons tenus pour vrais et qui, après révisions et alors que nous n'y adhérons plus, sont présents à notre esprit de manière différente – dire ici que c'est de manière « moins vive » est la façon la plus commode de noter cette différence. Inversement, si l'on peut établir qu'existe une différence de couleur doxastique dans la présentation des deux contenus, alors on peut présumer que celle-ci peut être attribuée à une différence dans le degré de l'adhésion de différents sujets à un même contenu ou d'un même sujet modifiant ses croyances dans le temps. Mais on peut aussi voir les choses de manière différentes et envisager un signal, toujours le même, qui accompagnerait les contenus d'adhésion, comme une petite lumière présente ici dans le cas du pessimiste qui pense que l'être humain est cupide et absente dans le cas où la même pensée serait envisagée par un esprit plus optimiste. L'adhésion serait alors un particulier mental que l'agent retrouve et reconnaît à l'occasion de différentes présentations et non une propriété affectant les contenus – un universel ou un déterminable – comme le suggère la métaphore de la couleur. Il est difficile de choisir entre ces deux options qui, bien qu'apparemment contradictoires, sont l'une et l'autre phénoménologiquement attractives.

Si l'adhésion vient par degrés, alors on a un argument en faveur de l'adhésion-couleur. Parmi les contenus auxquels nous adhérons, certains font l'objet d'une simple attente, d'autres d'une croyance, d'autres encore d'un savoir véritable. Dans chacun de ces cas nous nous sentons engagés de manière plus ou moins forte. « Je sais qu'il viendra » implique que j'adhère sans réserve au contenu « il viendra ». « Je suis persuadé qu'il viendra » ou « J'attends qu'il vienne » exprime un degré d'adhésion plus faible. Nos croyances se caractérisent par le degré de confiance qui leur est associé et, de manière dynamique, par leur capacité à résister au changement. Le degré d'adhésion représente le côté subjectif de chacun de ces aspects. Or si l'adhésion fonctionne comme un signal, il est plus difficile de rendre compte d'une gradation subjective, à moins de supposer qu'un signal différent dans chaque cas pourrait indiquer le degré de l'adhésion en même temps qu'il devrait l'attester. Mais ce serait alors faire du signal un déterminable mental et le défaire par là de son statut de particulier.

Je propose de garder en mémoire cette dualité phénoménologique. Quant à savoir ce qu'on peut en faire ou en tirer, c'est en effet une autre affaire et il est urgent de ne pas conclure trop vite. En particulier, nous n'avons pas de

raison de supposer que la différence dans l'adhésion est toujours une fonction de la justification. Il y a des cas dans lesquels la distribution *attendue* des degrés d'adhésion est manifestement fautive : il arrive que des croyances fondées sur des raisons très faibles soient plus ancrées, ou que des raisons faibles coexistent avec des raisons fortes mais soient davantage responsables de l'adhésion. Ainsi notre sociologue qui croit que l'être humain est cupide à la suite d'un événement caché dans les profondeurs de sa biographie, bien qu'elle connaisse les résultats des travaux sur la mesure de l'envie : elle voit les raisons, et les approuve, mais suit, pour ainsi dire, ses affections, et ce d'autant plus qu'elle peut le faire en partie à son insu s'il se trouve que raisons et inclinations tirent comme ici du même côté. Dans cette configuration, un particulier mental accompagne certaines représentations et les relie à l'agent d'une manière spécifique pour en faire des contenus de croyance sans qu'il soit véritablement possible d'y faire voir des nuances ou des degrés ; pourtant ces nuances existent certainement et une connaissance fine de la vie mentale du sujet permettrait assurément de les restituer. D'une manière plus générale si l'on doit distinguer, parmi les raisons, celles qui sont plus *responsables* de l'adhésion, autrement dit celles en vertu desquelles le sujet est prêt à asserter tels et tels contenus et à agir conformément à ces assertions, nous n'avons pas à supposer que les raisons les plus fortes du point de vue de l'explication sont également davantage source d'adhésion ou plus présentes à l'origine de nos convictions. Il est souhaitable qu'il en soit ainsi ; c'est sans doute un des réquisits de la rationalité ; mais c'est malheureusement loin d'être le cas le plus souvent.

On peut néanmoins penser que la dualité phénoménologique de l'adhésion correspond à une dualité plus profonde au niveau des attitudes épistémiques. Il est entendu et largement admis que la connaissance est factive – S sait que p entraîne que p est vrai – et que la connaissance, au moins la connaissance empirique, est également faillible : S peut savoir empiriquement que p sur la base de raisons que S pourrait également avoir si p était faux⁸. Il est vraisemblable que les contenus que nous tenons pour des connaissances sont identifiés comme tels et accompagnés d'un signal mental qui en marque la factivité, alors que ces mêmes contenus peuvent être éclairés de manière différente selon les raisons qui les soutiennent et donc sujets à différentes adhésions-couleurs.

Envisageons certains des exemples qui ont été imaginés par les épistémologues à l'occasion du débat autour de la relation de la croyance à la connaissance. Selon les uns, la croyance est une espèce de la connaissance, selon les autres, la connaissance vient en premier. Difficile de dire si l'une ou l'autre de ces options l'a d'ores et déjà emporté ou l'emportera dans l'avenir, mais le débat a au moins permis d'attirer l'attention sur la complexité des relations des deux genres d'attitude et sur l'existence de ces

8. Sur cette approche du faillibilisme et les précisions qu'elle requiert, voir R. Neta, « A Refutation of Cartesian Faillibilism », *Noûs*, 45/4, 2011, pp. 658-695, pp. 666-669.

cas où le sujet incontestablement sait que p mais n'est pas au clair avec sa croyance⁹. On peut savoir que p et ne pas croire que p – comme le mari trompé à qui on n'a presque rien caché mais qui s'entête –, savoir que p et croire que *non-p* – lorsqu'un résultat solide mais inattendu ne modifie pas immédiatement nos croyances plus anciennes – savoir et croire qu'on ne sait pas – comme l'étudiant imaginé par Radford qui pense ne rien savoir de l'histoire britannique et donne néanmoins de bonnes réponses lorsqu'on le questionne, par exemple, sur la date de la mort de la reine Élisabeth 1^{re}. Supposons que le mari trompé sait ; il dispose de tous les indices et sa femme a même abordé le sujet un soir, puis fait machine arrière lorsqu'il a coupé court par une plaisanterie. Il ne veut ou ne peut pas y croire. Il sait sans savoir qu'il sait. Plus tard, lorsqu'il saura complètement, les mêmes indices et contenus seront accompagnés du signal de l'adhésion épistémique. Ce sont pour le moment d'autres pensées qui se présentent à lui dans les couleurs de la croyance, par exemple que cet ami, qui a cherché à l'avertir, est décidément trop jaloux de sa réussite. Quant à l'étudiant de Radford qui donne correctement la date de la mort d'Élisabeth et pense qu'il ne la connaît pas : il adhère au contenu [Élisabeth est morte en 1603] seulement jusqu'à un certain point ou selon un certain degré. Mais aucune adhésion-signal ne l'invite à penser que la réponse qu'il a donnée a bien sa source dans le stock de ses connaissances véritables.

S'il est commode de décrire l'adhésion dans ces exemples selon deux modalités différentes, c'est surtout en raison de la dimension méta-cognitive qui caractérise les situations épistémiques respectives. Le mari trompé pense que les pensées de son ami bienveillant sont l'effet de sa jalousie et que les pensées qui le traversent lui-même et lui représentent sa femme comme infidèle ne sont pas justifiées. Ses croyances fausses portent sur d'autres croyances, les siennes propres ou celles d'autrui. Elles ressemblent plutôt à des désirs et partagent au moins avec ceux-ci la capacité d'augmenter et d'orienter l'attention. Elles n'annulent toutefois pas les indices et témoignages dont notre agent dispose et qui suffisent pour pouvoir lui attribuer un savoir de premier degré : il sait ce qu'il en est de son couple. De même pour l'étudiant de Radford qui pense que sa pensée [que la reine Élisabeth est morte en 1603] est l'effet d'un souvenir vague et probablement imprécis. Il pense qu'il pourrait avoir cette pensée même si elle était fausse et constate qu'elle n'est pas accompagnée du particulier mental qui signale les connaissances. Au moins pour le moment : on peut supposer qu'une fois le test effectué il modifiera ses vues sur ce point. Mais même au moment où il donne cette réponse (vraie) nous pouvons lui attribuer une connaissance qu'il ne s'attribue pas lui-même.

9. L'exemple séminal est en général attribué à C. Radford, « Knowledge—By Examples », *Analysis*, 27, 1966, pp. 1-11. Voir les exemples soumis à l'approche expérimentale dans B. Myers-Schulz et E. Schwitzgebel, « Knowing that P without Believing that P », *Notus*, 47/2, 2013, pp. 371-384, que je reprends pour l'essentiel.

Il adhère selon un certain degré au contenu [Élisabeth est morte en 1603] mais il n'y adhère pas selon l'adhésion-signal.

4. Allégation et auto-attribution

Il suit des exemples évoqués jusqu'ici que la métaphore de l'adhésion s'applique dans des contextes d'attitudes variées dont ce qu'on appelle les « conditions de satisfaction » sont différentes. Une *croissance* représente le monde ou une partie du monde d'une manière déterminée et exclut les versions du monde incompatibles avec son contenu. Si je crois qu'il pleut, toutes les versions du monde dans lesquelles il ne pleut pas autour du lieu où je me trouve sont rejetées ou considérées par moi comme inactuelles. Cette manière répandue et éclairante d'envisager les choses autorise l'utilisation de la machinerie sémantique des possibilités dans l'étude des modalités épistémiques¹⁰. Elle permet également d'établir des points de passage entre les pensées pratiques et les attitudes épistémiques. Si, par exemple, les versions du monde où il fait beau sont exclues de la classe de mes croyances – parce que je crois qu'il pleut – et si j'aime faire du vélo seulement quand il fait beau, les versions du monde où je fais du vélo sont également exclues, mais éventuellement cette fois des scénarios qui cartographient mes intentions. Les intentions et les croyances se rencontrent ainsi par leurs conditions de satisfaction.

Bien que ces conditions de satisfaction soient objectives, elles entretiennent avec l'adhésion un lien particulièrement manifeste dans le cas de la croyance. Un agent ne peut pas croire que *p* s'il ne tient pas *p* pour vrai et, inversement, ne peut pas tenir *p* pour vrai et ne pas croire que *p*. Or, que *j'adhère* à un contenu de croyance signifie au moins que je le tiens pour vrai. Ce sont les conditions de satisfaction qui confèrent ici un sens à la métaphore de l'adhésion, au point que celle-ci peut sembler une autre manière de désigner celle-là :

Adhésion₁ : le sujet *S* adhère au contenu de croyance *p* ssi les conditions de satisfaction de la croyance dont le contenu est *p* sont réunis pour *S*.

Peut-on généraliser ce point et conclure qu'« adhérer » est une expression métaphorique du « tenir pour vrai » ? La difficulté est ici que la métaphore s'applique non seulement à la croyance, mais aussi à des contextes d'attitudes pour lesquelles la vérité n'est pas normative comme elle l'est pour la

10. Voir par exemple D. Chalmers, *Constructing the World*, Oxford, Oxford University Press, 2012, pp. 233-244. Pour une approche différente, voir S. Yalcin, « Nonfactualism about Epistemic Modality », in *Epistemic Modality*, A. Egan et B. Weatherson (dirs.), Oxford, Oxford University Press, 2011, pp. 295-332.

croissance. Ne peut-on pas adhérer à une imagination, à une intention et, en général, à des contenus qu'on ne va pas chercher à déclarer vrais¹¹ ?

Lorsqu'on soutient que la vérité est normative relativement à la croyance, on fait référence au fait que si *p* fait partie des croyances d'un agent, nécessairement l'agent tient *p* pour vrai¹². Tout ce à quoi l'agent croit est dans l'extension de ce qu'elle tient pour vrai¹³. Ce n'est pas le cas, par exemple, pour l'imagination ou l'intention. Si l'agent imagine que *p*, alors elle ne tient pas *p* pour vrai mais envisage *p* comme vrai. Les conditions de vérité de *p* sont à l'imagination dont le contenu est *p* ce que la valeur de vérité de *p* est à la croyance de même contenu. Quant au désir et à l'intention, ils se situent, relativement à la vérité, encore un peu plus loin. Un désir de faire du vélo est satisfait lorsqu'il est réalisé, c'est-à-dire lorsque la situation dans laquelle le sujet fait du vélo est actuelle. Mais un agent peut désirer faire du vélo et savoir qu'elle n'en fera jamais – désirer que *p* et savoir que *p* est faux –, si les conditions de satisfaction de ce désir appartiennent à d'autres mondes possibles, par exemple aux mondes où elle ne vit pas dans une chaise roulante. Dans ce cas, faire du vélo *fait partie des désirs* de l'agent tout comme *p* – que l'agent tient pour vrai – *fait partie* de ses croyances. Mais cette propriété d'appartenance ne dépend alors, ni de la valeur de vérité de *p*, ni de ses conditions de vérité. S'il y a un sens à dire que l'agent « adhère à son désir de faire du vélo », c'est dans la mesure où elle se reconnaît ce désir, se l'auto-attribue, ou au moins le manifeste par son comportement. Dans le cas du désir, la condition de satisfaction n'est donc pas *ipso facto* une condition nécessaire de l'adhésion. Celle-ci porte plutôt sur la relation à l'agent, sur le degré auquel l'agent se reconnaît porteur d'un désir – ou, dans le cas plus difficile d'un désir inconscient, le degré auquel celui-ci pourrait lui être attribué.

Que doit-on conclure de ces comparaisons ? Que l'adhésion est un mixte de *tenir pour vrai* – allégation – et de *tenir pour sien* – auto-attribution. Dans les contextes d'intention et de désir, les deux aspects sont disjoints. Dans les contextes doxastiques au contraire, l'auto-attribution et l'allégation coïncident ou se recouvrent, de sorte que la métaphore de l'adhésion peut renvoyer confusément à l'une et à l'autre. Il est donc important pour nous d'envisager avec précision la nature de ce recouvrement. Il peut en effet s'agir

11. On considère en général que les conditions de satisfaction d'une croyance résident dans sa *valeur de vérité*, que les conditions de satisfaction d'un contenu imaginé sont ses *conditions de vérité* ; que les conditions de satisfaction d'une intention ou d'un désir sont des *conditions de réalisation* ; enfin qu'une perception est satisfaite si son contenu est non-illusoire, autrement dit selon qu'elle est ou non *véridique*.

12. On s'est beaucoup disputé sur ces aspects normatifs de la vérité, surtout dans le contexte du débat sur les conceptions minimales et déflationnistes. Voir P. Engel, « Is Truth a Norm ? », in P. Kotatko, P. Pagin et G. Segal (dirs.), *Interpreting Davidson*, Stanford, Center for the Study of Language and Information, 2001, pp. 37-51.

13. Si, comme l'affirment les partisans du « *knowledge first* », les connaissances ne sont pas des cas de croyances, alors on peut savoir que *p* sans tenir *p* pour vrai. Dans ce cas, le tenir pour vrai est plutôt l'expression d'un savoir de second ordre : je tiens pour vrai quand je sais que je sais.

d'un lien plus ou moins fort. Selon une conception extrême et dans le cas spécifique des croyances, le *tenir pour vrai* et le *tenir pour sien* seront tenus pour identiques. Dans le cadre d'une conception de ce genre, disons une conception *vériste*, la métaphore de l'adhésion n'apporte pas grand-chose, puisque le vérisme implique que tout ce qui peut se dire de la croyance en termes d'adhésion peut également se dire au moyen de l'allégation. Mais la relation d'identité n'est-elle pas ici trop forte ?

Une des questions soulevée par l'absurdité des assertions mooréennes porte exactement sur la nature du recouvrement envisagé ici :

Absurdité De Moore (Ma) : les assertions de phrases de l'une et l'autre de ces deux formes sont absurdes :
 p et je ne crois pas que p
 p et je crois que *non- p*

La plupart des explications de MA s'appuient sur la règle – souvent attribuée à Grice¹⁴ – selon laquelle il est requis, pour communiquer un contenu [p] au moyen d'une assertion de p , de se présenter soi-même comme croyant que p . Celui ou celle qui affirme p manifeste l'intention de communiquer p et manifeste l'intention contraire en déclarant « je ne crois pas que p ». Il s'agit donc d'une contradiction qui se situe au niveau de l'agir. MA n'implique pas que les contenus assertés dans :

- A_1 « B est une croyance vraie »
 et
 A_2 « B est une de mes croyances »

ont la même signification. Il existe de nombreuses situations qui attestent du contraire.

Supposons que D. pense être assis au coin du feu, alors qu'il s'est également convaincu que toutes ses croyances sont sujettes à quelque doute et qu'aucune ne peut à proprement parler être tenue pour assurée. On ne rend pas compte complètement de cette situation en notant que D. croit que sa pensée [qu'il est assis au coin du feu] est – peut-être – fausse ou croit qu'il *n'est pas* présentement au coin du feu. Ce que pense D. est plutôt : que, compte tenu de ses raisons de douter, cette pensée ne fait pas partie de ses croyances *contrairement à ce qu'il croyait*. Autrement dit : D. se représente l'ensemble de ses croyances et constate qu'il contient un intrus, une pensée qui ne peut pas se trouver là puisqu'elle est (possiblement) fausse : une pensée qui, ne faisant plus l'objet d'une allégation suffisante, doit être déplacée en dehors de la sphère de l'auto-attribution. Il est clair qu'un état mental de ce genre instancie une cognition étagée, dans laquelle l'allégation et l'auto-attribution n'interviennent pas au même titre : cette pensée que D. se reconnaît encore comme

14. Voir T. Chan « Belief, assertion and Moore's Paradox », *Philosophical Studies*, 2008, vol. 139 (3), pp. 395-414 à qui j'emprunte la formulation de MA.

une croyance au titre de l'auto-attribution, il ne la reconnaît pas comme telle au titre de l'allégation. Et dans le cas où le doute est assez puissant, le sujet pense qu'il peut réitérer cette métareprésentation pour chacun de ses contenus doxastiques, et se représente alors l'ensemble de ses croyances en tant qu'il ne contient que des intrus. Si A_1 et A_2 avaient la même signification, une pensée dont le contenu serait la métacroyance que telle croyance de premier ordre est fausse et une pensée constatant simplement l'absence de cette croyance seraient indiscernables, de même que seraient indiscernables l'état d'un sujet qui doute – lorsque ce doute porte sur la quasi-totalité de son stock doxastique – et l'état d'un sujet dont le stock doxastique est presque vide et qui, pour cette raison, n'a probablement pas même accès aux raisons de douter.

A_1 et A_2 ne sont donc pas synonymes. Il est vrai néanmoins qu'elles peuvent se substituer l'une à l'autre dans la plupart des situations, et qu'elles permettent de faire l'une et l'autre dans ces situations à peu près la même chose avec le langage : elles sont, en ce sens, des paraphrases¹⁵. Mais cette relation – être des paraphrases – qualifie la classe des situations où elle s'applique ; elle requiert, par exemple, des conditions minimales de rationalité sur les agents. Comme ces conditions ne sont pas toujours remplies et viennent par degrés, il est important de distinguer les deux aspects de la croyance qui s'expriment respectivement dans A_1 et A_2 .

Les croyances que nous pensons pleinement justifiées font partie de nos croyances, à nos propres yeux, en vertu de leur justification. Mais certaines de nos croyances solides sont faiblement justifiées et, lorsque nous en sommes conscients, ce qui est solide alors, c'est plutôt le lien qu'elles entretiennent à nous-mêmes, c'est-à-dire l'auto-attribution dont elles font l'objet. On peut supposer par exemple que, pour une bonne partie des savants d'aujourd'hui qui ont des croyances religieuses mais ne croient pas, par exemple, que l'âge de la terre est plus ou moins 6 000 ans, les croyances religieuses sont faiblement justifiées, mais sans doute fortement auto-attribuées si elles font l'objet d'une adhésion. Le second sens de l'adhésion qui se fait jour ici peut être précisé ainsi :

Adhésion₂ : le sujet S adhère au contenu p si (i) S ne peut se représenter p sans se l'attribuer comme une de ses croyances et/ou (ii) S ne peut se représenter lui-même autrement que comme entretenant la croyance que p .

Adhésion₂ est une fonction du sujet à son propre stock doxastique. Elle associe à chaque croyance la propriété de faire ou de ne pas faire partie du stock et à chaque représentation du stock la propriété d'être ou non le stock de tel sujet. Les deux conditions (i) et (ii) ainsi que leur relation portent donc

15. Sur cette conception de la paraphrase, cf. J.-B. Rauzy, « Présomptivisme », in Ph. Büttgen et J.-B. Rauzy, (dirs.), *La Longue Durée, Pour Jean-François Courtine*, Paris, Vrin, pp. 313-336.

sur le profil modal de l'auto-attribution. Avant de s'y arrêter plus longuement, il convient d'envisager une objection très générale à cette version de l'adhésion.

On peut en effet à nouveau insister sur la dimension du désir et objecter que les états mentaux qui font l'objet d'Adhésion₂ ne sont pas à proprement parler des croyances. Nous adhérons à certains contenus en vertu des raisons que nous avons de les tenir pour vrais et ces contenus pris ensemble forment notre stock doxastique. Mais nous adhérons aussi à d'autres contenus d'une manière différente : nous reconnaissons ces contenus comme nôtres indépendamment des justifications dont nous disposons à leur sujet. Nous les traitons alors comme s'ils n'étaient pas des porteurs de vérité, comme des désirs ou des inclinations, bien qu'ils soient effectivement aptes au vrai. Ce sont les contenus de ce genre qui peuvent faire l'objet d'Adhésion₂ : ils constituent éventuellement des quasi-croyances, dit l'objection, mais certainement pas des croyances authentiques.

Les attitudes non-cognitives, largement étudiées dans le champ de la philosophie morale, peuvent servir ici de point d'appui. Un désaccord est l'effet d'une attitude non-cognitive lorsqu'il n'est pas le résultat d'une différence dans la situation informationnelle des agents et ne peut être résolu par la seule modification de cette situation. Si S croit qu'une action est bonne, quand S' pense le contraire et si, en outre, aucune augmentation du savoir de S et de S' n'est de nature à rapprocher leurs vues, alors le désaccord de S et de S' doit être conçu en dehors du champ doxastique proprement dit. Les auteurs qui insistent sur cette dimension non-cognitive sont effectivement conduits à concevoir l'adhésion à un contenu moral (ou *moral acceptance*) comme une forme du désir : un désir qui s'accompagne de la tendance à concentrer l'attention sur l'objet du désir et à le présenter sous un jour favorable¹⁶. Et ce désir n'est pas sensible à la justification et au tenir pour vrai. Telle est alors notre objection : si, comme on en convient le plus souvent, les attitudes morales, ou au moins celles qui peuvent être conçues de cette manière, ne sont pas des croyances, alors nous n'avons pas de raison de penser que les contenus non-moraux qui peuvent être l'objet d'une adhésion au sens d'Adhésion₂ le sont davantage. Une conviction religieuse, par exemple, peut être alimentée, en dépit des invraisemblances qu'elle contient, par la simple représentation de son objet, donnant lieu à cette adhésion non-cognitive qui signale une attitude non-doxastique. Selon cette ligne, on distinguera les croyances comme les attitudes pour lesquelles le tenir pour vrai intervient dans le tenir pour sien, – autrement dit : les attitudes pour lesquelles l'auto-attribution est une fonction de la justification – et les

16. C'est la position défendue par M. E. Kalderon dans *Moral Fictionalism*, Oxford, Oxford University Press, 2005, p. 45 suivant T. Scanlon, *What we Owe to Each Other*, Cambridge, Mass, Belknap Press, 1998, chap. 1. Voir aussi M. E. Kalderon, « Moral Fictionalism, the Frege-Geach Problem and Reasonable Inference », *Analysis*, 68/2, 2008, pp. 133-143.

autres attitudes – dont fait éventuellement partie l'acceptation morale – qui appartiennent généralement à la famille des désirs.

En dépit du consensus dont elle fait l'objet, il y a néanmoins quelque chose dont cette objection et le tableau tranché qu'elle impose ne parviennent pas à rendre compte. Lorsque nous adhérons à un contenu sur la base d'Adhésion₂, nous savons que le contenu en question est susceptible d'être vrai et nous pensons entretenir rien de moins qu'une croyance. Rien, dans la phénoménologie des états mentaux, ne nous invite à penser que nous aurions un dispositif propre à modifier la segmentation des contenus lorsque nous adhérons de cette manière. En ce sens les deux formes de l'adhésion décrivent bien deux relations à la croyance. Revenons à cet esprit religieux qui, comme beaucoup je suppose le font, admet que l'âge de la terre n'est pas plus ou moins 6 000 ans. Il serait embarrassé si, fort de l'analyse précédente, un psychologue le portracterait en soutenant qu'il faut, d'un côté, ranger ses croyances authentiques – parmi lesquelles ses croyances relatives à l'âge de la terre – et, d'un autre côté, des quasi-croyances auxquelles il adhère plutôt par la représentation qu'il a de lui-même – parmi lesquelles figureraient les bases de ses principales convictions religieuses. Il serait embarrassé, non pas parce qu'il serait nécessairement fermé à toute distinction de ce genre ou à toute distinction dans les modes de ses justifications. S'il admet que, sur l'âge de la terre, il ne faut pas faire comme Bellarmin¹⁷ et tout prendre à la source révélée, on peut supposer qu'il est capable, par généralisation, de distinguer ses croyances bellarminiennes de celles qui ne le sont pas. Il serait plutôt embarrassé parce que ses croyances bellarminiennes sont également pour lui *des croyances* au sens précis où il les tient pour vraies. Dans le miroir psychologique qu'on lui tend, elles apparaîtraient sous un autre registre, en compagnie des intentions et des désirs, et il y a fort à parier que le stock doxastique qui ferait alors la teneur du portrait serait à ses yeux illégitimement amputé. Adhésion₂ s'appuie sur la dimension épistémique de la subjectivité elle-même. Le stock de ce qui est tenu pour vrai par un agent à un certain moment et dans le temps caractérise l'agent. La subjectivité désigne au moins autant une région dans l'espace épistémique qu'un usage de la première personne ou une série d'expériences phénoménales.

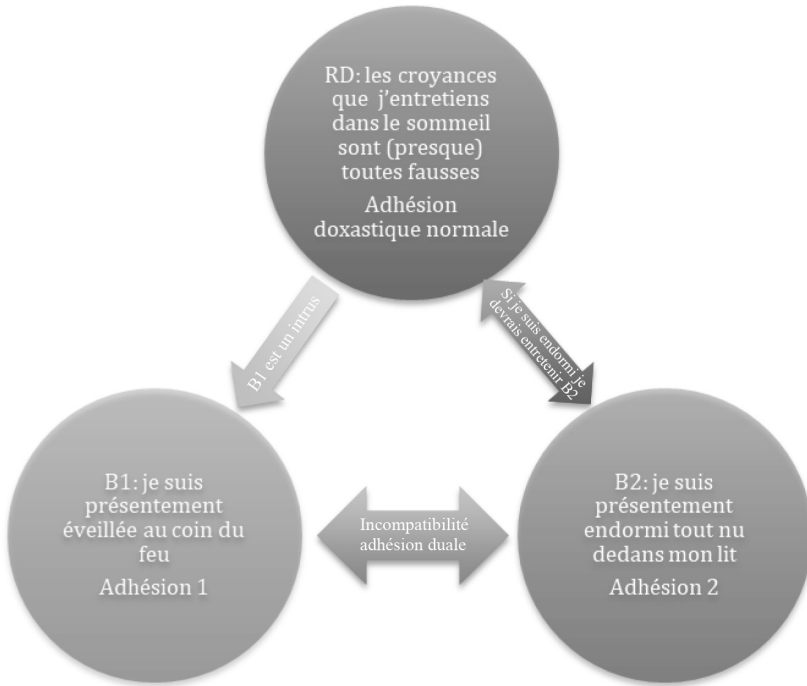
Considérons à présent le cas de notre méditant qui est assis au coin du feu et doit réviser ses croyances en fonction de l'effet qu'il est prêt à attribuer à ses « raisons de douter ». Les deux formes de l'adhésion me semblent

17. Je reprends ici l'exemple heureux de Boghossian. Le cardinal Bellarmin a mené l'accusation dans le procès de Galilée. Boghossian le décrit comme un esprit rationnel, qui refuse le genre d'évidences alléguées par Galilée lorsqu'il s'agit de dessiner la carte du ciel, puisqu'il y a des informations relatives à la carte du ciel dans la Bible : on peut dire que Bellarmin applique un ordre épistémique lexical selon lequel les justifications prises à la source révélée l'emportent sur toutes les autres lorsqu'elles existent. Mes croyances bellarminiennes sont celles qui respectent cet ordre. Voir P. Boghossian, *Fear of Knowledge. Against Relativism and Constructivism*, Oxford, Clarendon Press, 2006, pp. 59 et 103.

également impliquées. Dans cette situation en effet, celui que nous avons appelé D. pense que les croyances qui ont été jusqu'ici les siennes sont entachées de doutes, y compris la croyance selon laquelle il est présentement au coin du feu. Compte tenu de ses raisons de douter, cette pensée [qu'il est présentement au coin du feu] – disons B_1 – ne fait pas partie de ses croyances *contrairement à ce qu'il a pu croire* lorsqu'il n'avait pas encore envisagé les choses de cette manière, c'est-à-dire lorsqu'il n'avait pas encore accès aux raisons de douter. Tout en demeurant inchangée quant à son contenu, B_1 est devenu un intrus à la suite de la modification du réglage épistémique, depuis que le type de la justification ordinaire a cédé à l'exigence de justification exceptionnelle propre au contexte sceptique. Il me semble très contestable de décrire une situation de ce genre en avançant simplement que les conditions de satisfaction de la croyance B_1 ne sont plus réunies pour D. et que, par suite, D. n'y adhère plus, au sens d'Adhésion₁. Car D. ne cesse pas de tenir B_1 pour vraie ; il découvre aussi qu'il ne la tient pas comme assurée. Il est clair que, du point de vue de D., ce qui nous apparaît comme un réglage épistémique requiert l'intervention de croyances nouvelles, croyances qui font elles-mêmes l'objet d'une adhésion. Ainsi D. tient pour vraie la raison de douter (RD) selon laquelle la croyance [qu'il est présentement assis au coin du feu] (B_1) est possiblement fausse. (RD) est une métacroyance. Par exemple la croyance selon laquelle il existe des conditions exceptionnelles dans lesquelles les croyances du type B_1 se révèlent fausses, conditions qui ne sont peut-être pas réunies, mais que rien n'exclut complètement.

Descartes insistait sur le fait que la méditation devait être effectivement exercée et non simplement rapportée ou décrite. Si la métacroyance intervient comme telle dans le compte rendu de l'expérience de D., dans l'expérience elle-même, elle prend aussi la forme d'une croyance de premier ordre (B_2) : par exemple la croyance exprimée par « je suis présentement endormi ». Le récit de la révision effectuée par D. doit faire voir que l'adhésion à RD désactive l'adhésion à B_1 et, en général, l'adhésion à toutes les croyances de premier ordre visées par cette raison de douter, c'est-à-dire l'adhésion aux croyances qui figurent dans le champ de son *aboutness*. Mais il doit également faire justice à cette dimension d'exercice, par quoi RD est instanciée et intervient dans sa liaison à une pensée actuelle de premier ordre. Dans le compte rendu de la méditation devront donc intervenir successivement B_1 (la croyance de premier ordre qui fait l'objet de la révision) ; RD : la croyance d'ordre deux qui porte (entre autres) sur B_1 et établit qu'elle est douteuse ; enfin B_2 : une autre croyance de premier ordre incompatible avec B_1 à laquelle le sujet adhère d'une façon différente, ce qu'exprime la forme suppositionnelle du « supposons à présent que je suis endormi. » Bien qu'elles soient logiquement du même ordre, B_1 et B_2 sont nécessairement différentes d'un point de vue phénoménal et l'une semble pour ainsi dire forcée. D. suppose – mais le croit-il ? – qu'il est à présent endormi : il introduit cette croyance d'ordre 1 dans l'extension

d'Adhésion₂. On peut rendre compte de cette différence par la forme duale de l'adhésion comme dans la figure suivante :



On peut donc adhérer à un contenu selon l'une ou l'autre des deux formes et souvent, au moins dans les situations concrètes de révision, les deux formes de l'adhésion sont également impliquées.

5. Justificationnisme

La dimension de l'adhésion est un trait commun aux attitudes doxastiques et aux attitudes épistémiques. C'est une des raisons pour lesquelles tant d'auteurs, dans la tradition et récemment, ont fait de la connaissance une espèce de la croyance. Un sujet peut avoir accès à une forte justification en faveur d'un contenu, mais s'il n'y adhère pas, on ne dira pas qu'il s'agit d'une de ses connaissances. Il est probable, pour reprendre encore l'exemple de Boghossian, que le cardinal Bellarmin comprenait les arguments de Galilée en faveur du copernicanisme et qu'il avait les capacités de regarder dans la lunette astronomique, d'y faire des observations et de réeffectuer les calculs de géométrie céleste nécessaires. On peut supposer qu'il l'a fait. Mais il n'a pas pour autant adhéré au copernicanisme et donc ne reconnaissait pas au copernicanisme le statut d'une connaissance. On peut avoir *pleinement*

accès à une justification sans adhérer au contenu de pensée auquel la justification s'applique. Cela signifie-t-il que l'adhésion n'est *en rien* une fonction de la justification ? Bellarmin lui-même ne disposait-il pas d'un autre type de justification auquel il pouvait adosser ses croyances à propos des orbites des planètes ? Dans le but d'aborder cette question plus en détail, convenons d'appeler *justificationnisme* la thèse selon laquelle l'adhésion à un contenu est toujours le résultat ou l'effet d'une justification : pour tout contenu auquel un sujet adhère, il existe une justification – active pour le sujet – qui produit l'adhésion et l'emporte sur les autres justifications.

Justificationnisme : le sujet S adhère au contenu p ssi (i) il existe une justification J pour p , (ii) J est active en S et (iii) J est plus forte que les justifications de *non-p* qui sont également actives en S.

On objectera aussitôt qu'une « justification active en S » désigne sans aucun doute un contenu mental de S, un contenu de croyance qui entre en relation à un autre contenu, mais donc un contenu auquel S adhère, de sorte que la condition (ii) revient à dire que S adhère à un contenu s'il existe un contenu auquel S adhère. N'y a-t-il pas un vice dans la définition ? Ou, au moins, une régression ? On sait qu'on a affaire à une régression analogue lorsqu'on précise les conditions de la relation à la justification elle-même – la justification d'une croyance est une croyance qui, le plus souvent, est adossée à une justification – et qu'on n'a pas pour autant renoncé à en faire usage¹⁸. L'important est plutôt le point suivant : pour la justificationniste, l'adhésion de Bellarmin à tel système du monde n'est pas quelque chose de différent, ou quelque chose de plus, que le corps des croyances qui valent justifications et ont pour lui une force épistémique supérieure dans le domaine envisagé¹⁹. Autrement dit : si le justificationnisme est vrai, alors le projet d'une enquête sur l'adhésion paraît assez vain et le gain qu'on doit en attendre pour l'épistémologie très faible. L'essentiel de mon effort dans ce qui suit consiste à amender le justificationnisme ainsi conçu pour l'adapter autant que possible aux cas difficiles, c'est-à-dire au cas dans lesquels l'adhésion semble indépendante ou liée à d'autres dimensions de la vie doxastique. Je pense que le justificationnisme doit être défendu dans une certaine mesure ; qu'il est *un préférable* et caractérise les situations épistémiques les plus favorables ; que, à certains égards, une *bonne* situation épistémique est une situation justificationniste. Par suite, il est important de réfléchir, par

18. Voir D. M. Armstrong, *Belief, Truth and Knowledge*, Cambridge, Cambridge University Press, 1973, p. 153 : « [...] the threatened infinite regress—first notice in the literature by Plato at the end of the *Theaetetus* (209^E-210^B)—does not leave the supporter of an Evidence-condition completely without resource ».

19. Il est assez naturel de supposer que les types de justification actifs dépendent d'éléments contextuels et sont en outre indicés sur les domaines ou « sujets » de nos croyances. Pour l'élaboration de la notion, qui s'avère délicate, voir S. Yablo, *Aboutness*, Princeton, Princeton University Press, 2014.

des considérations sur les aspects métacognitifs, collectifs et phénoménaux des croyances justifiées, aux conditions dans lesquelles une situation épistémique donnée peut être rendue justificationniste.

Revenons à notre chercheuse, convaincue que l'être humain est cupide ; on peut l'imaginer engagée dans un échange conversationnel privé ou dans une discussion savante à l'occasion d'un colloque. Le même contenu de croyance [que l'être humain est cupide] sera présenté différemment dans l'un et l'autre cas, soutenu par des arguments différents et orienté vers un but propre : éveiller, par exemple, la compassion d'un juge d'appel dans le contexte de son divorce, promouvoir une forme de la coopération sociale dans le cadre d'une communication académique. Il est probable que, dans la situation savante, elle évitera de faire part de ses déconvenues privées pour présenter sa conviction, non pas parce que, d'une manière générale, on ne fait pas part de ses déconvenues privées dans un colloque – il existe des colloques où c'est au contraire bien venu – mais plutôt parce qu'il est essentiel au type d'adhésion qu'elle cherche à obtenir qu'il soit le résultat d'une justification rationnelle élaborée et qu'il est de connaissance commune – dans la situation épistémique du colloque – que celle-ci ne peut se fonder sur une expérience singulière et sur un ensemble d'évidences²⁰ tronquées. Nous savons que l'adhésion à un contenu n'est pas nécessairement le résultat d'une bonne justification, que l'adhésion produite par une justification plus solide n'est pas elle-même plus solide : c'est la raison pour laquelle nous nous engageons conjointement dans des activités de connaissances et dans des échanges savants sur la base de règles métacognitives partagées. Le lien de l'adhésion à la justification est donc en partie au moins *institué* par ces règles qui, sous la forme d'une information accessible d'ordre élevé, contribuent à la situation épistémique.

La manière la plus simple de rejeter le justificationnisme consiste à établir en général la possibilité de l'indifférence épistémique. Nous sommes tous d'accord pour dire que l'adhésion est sensible à la justification, au sens où un contenu qui est justifié d'une manière ou d'une autre est plus susceptible d'être cru par celui ou celle qui a accès à cette justification qu'un contenu qui ne l'est pas. Mais, encore une fois, cela ne signifie pas que l'adhésion est une simple fonction de la justification. Une justification solide ne suffit pas toujours à détruire la croyance du contraire et il arrive que des croyances antagonistes survivent dans le même agent alors que leurs justifications sont très inégales. Un esprit qui fait preuve d'indifférence épistémique est un esprit qui a accès aux raisons, qui maîtrise les règles d'inférence, qui manipule correctement l'inférence matérielle et qui néanmoins est capable de ne pas adhérer à un contenu auquel aurait dû le conduire la manipulation des raisons. Jusqu'à où l'indifférence épistémique est-elle possible ?

20. Ici et dans la suite, j'emploie le français « évidence » comme un néologisme pour l'anglais « evidence » dont l'usage est assez fréquent en épistémologie pour qu'on se décide à éviter les périphrases.

Essayons de construire un argument de concevabilité dans le style de Chalmers²¹. Un zombie philosophique est un être imaginaire qui nous ressemble du point de vue des propriétés physiques mais qui n'a pas de conscience phénoménale. Appelons de même « zombie épistémique » un être imaginaire qui nous ressemble du point de vue des propriétés physiques, mais qui n'a pas de sensibilité épistémique, est indifférent aux raisons, et même *manifeste une indifférence épistémique maximale*. Encore une fois, le zombie épistémique n'est pas dépourvu de rationalité. On peut même, pour la force de l'argument, supposer qu'il y a autant de diversité dans les zombies épistémiques qu'il y en a dans les êtres humains sous le rapport de la rationalité : certains y ont un accès laborieux, d'autres sont au contraire des experts dans l'invention et le maniement des raisons. Les experts peuvent néanmoins être des zombies s'ils sont capables d'inventer des raisons très fortes, de comprendre en quoi ce sont des raisons, mais se sentent libres de ne pas y adhérer et de ne pas agir selon ces raisons, ou de ne pas inférer selon ces raisons. L'argument à la Chalmers serait le suivant :

1. Les zombies épistémiques sont concevables.
2. Si les zombies épistémiques sont concevables, alors ils sont possibles.
3. Si les zombies épistémiques sont possibles alors l'adhésion n'est pas une fonction simple de la justification et le justificationnisme est faux.

La difficulté principale se situe dans la première prémisse. Avancer que les zombies épistémiques sont concevables, c'est en effet faire appel à une notion d'adhésion séparée de la justification, et par là à quelque chose que l'argument dans son ensemble doit lui-même établir. En effet, si le zombie épistémique est rationnel, on peut supposer qu'il est sensible à la relation :

$$R(B_1, B_2) : B_1 \text{ est la raison de } B_2$$

où B_1 et B_2 sont des croyances. Si cette relation n'existait pas pour le zombie, alors cela n'aurait pas beaucoup de sens de maintenir qu'il est rationnel : il ne pourrait pas, par exemple, inférer la fausseté d'un contenu à partir d'une contradiction. Notre zombie est donc capable d'adhérer à B_1 , de savoir que B_1 est la raison de B_2 , de savoir même que c'est une raison très forte de B_2 et néanmoins de ne pas adhérer à B_2 . Il est clair que le zombie épistémique n'est pas concevable de cette manière. Soit $[B_2]$ le contenu propositionnel de B_2 . Un zombie épistémique serait précisément défini comme un sujet rationnel capable d'asserter :

(PM) Je crois que $[B_2]$ (en vertu de B_1) et non- $[B_2]$ (je n'y adhère pas)

qui est une version de l'absurdité de Moore. Faut-il alors considérer que l'inconcevabilité des zombies épistémiques est établie et que le

21. D. J. Chalmers, « Consciousness and its Place in Nature », in *The Character of Consciousness*, New York/Oxford, Oxford University Press, 2010, pp. 103-139.

justificationnisme a marqué un point décisif ? Il est encore possible d'établir que les zombies épistémiques sont concevables de manière contextuelle, ou à la *Descartes*.

Supposons un agent S dont on sait qu'elle est un être humain rationnel mais qu'on se propose de transformer autant que possible en un zombie épistémique. (i) On parvient tout d'abord à convaincre S que la relation « ... est la raison de... » est absolument univoque, que c'est une relation qui n'admet pas de degrés, de sorte que si quelque chose s'y oppose, quoi que ce soit, alors la relation tombe. *La relation R (être la raison de) ne survit pas à la plus petite incertitude* (il se peut que S résiste sur ce point et il faudra alors la convaincre d'admettre ce principe de la plus petite incertitude de temps de l'expérience). (ii) Deuxièmement, on soumet à S une croyance B_1 à laquelle S adhère : B_1 fait partie des croyances de S . On demande à S une justification pour B_1 , ou, si l'on veut, puisque nous ne croyons pas dans l'indifférence épistémique, la raison de son adhésion à B_1 . S donne une raison B_2 . Si B_1 a été choisie assez judicieusement, non seulement B_2 est telle que $R(B_2, B_1)$ pour S , mais S sait que les croyances du genre de B_2 valent en général comme raison des croyances du genre de B_1 . Autrement dit : elle sait où aller chercher sa raison et ce « où aller chercher » est au moins l'embryon d'une justification d'ordre élevé. La raison pour laquelle B_2 vaut comme raison de B_1 réside dans la relation (de raison) entre le genre de B_2 et celui de B_1 . Par exemple la raison pour laquelle S est justifiée à croire que l'objet placé devant est bien rectangulaire comme elle le perçoit est que les données de la perception (genre de B_2) sont en général fiables dans les conditions de la perception présente et justifient les croyances perceptuelles (genre de B_1). (iii) Troisièmement : on introduit une nouvelle relation entre des croyances, dérivée de R : la raison de douter.

$RD(B_1, B_2)$: B_1 est la raison de douter de B_2

Dans le présent contexte, la relation R (être la raison de) ne survit pas à la plus petite incertitude : B_1 est la raison de douter de B_2 signifie que la croyance B_1 annule la classe des croyances B_n qui sont avec B_2 dans la relation R . On soumet à S une raison de douter de B_1 $RD(B_1)$. Il n'est pas très difficile de le faire puisque S est devenue très sensible à ce genre de raison. Par exemple, si le contenu de B_1 est perceptif, on fait remarquer à S que la perception n'est pas absolument fiable. De cette manière, au sortir de l'étape (iii) nous sommes dans une situation où S sait ce que c'est qu'une raison, S est même éventuellement une experte dans le maniement des raisons, S a une bonne raison pour B_1 (RB_1) mais S n'adhère pas à B_1 . C'est *en gros* la situation d'indifférence épistémique que nous recherchons.

Pourquoi est-ce *en gros* seulement la situation d'indifférence épistémique ? Premièrement nous sommes sous la dépendance de la première étape : il faut que nous ayons réussi à convaincre S qu'« être la raison de... » ne survit pas à la plus petite incertitude. Ceci suppose une élaboration contextuelle

importante. Mais on peut supposer que S est tentée de faire cette expérience, qu'elle est tentée de savoir ce que cela fait d'être un zombie épistémique et que, si elle pense qu'il faut pour cela accepter cette condition, alors elle l'acceptera. Deuxièmement, la situation où l'on obtient cette indifférence épistémique est elle-même épistémiquement instable : la raison de douter RDB_1 engendre une croyance plus forte B_2 , qui est soutenue par une raison RB_2 et qui requiert à son tour une raison de douter nouvelle (RDB_2) pour que la situation d'indifférence épistémique soit rétablie. Mais bien qu'elle soit instable, la situation dans laquelle on se trouve en présence de RDB_1 est bien une situation d'indifférence épistémique. Appelons-la, si l'on préfère, *indifférence épistémique fugitive*. N'a-t-on pas le droit de penser que le sujet S dans cette situation est un zombie épistémique, même fugitivement ? Qu'il existe un scénario dans lequel on peut reconnaître une situation de ce genre suffit pour notre première prémisse : puisque nous pouvons imaginer l'expérience d'une indifférence épistémique, et même faire cette expérience, il s'ensuit que les zombies épistémiques sont concevables et donc que l'adhésion n'est pas une fonction simple de la justification.

6. L'uniformité modale doxastique et la psychothérapie modale de Yablo

Nous avons les moyens conceptuels de comprendre pourquoi il en est ainsi. La raison pour laquelle l'adhésion ne saurait être une fonction simple de la justification réside dans sa forme duale d'allégation et d'auto-attribution. Une chose est de servir de fondement au tenir pour vrai – ce que fait principalement la justification – autre chose de provoquer le tenir pour sien – ce qu'elle fait aussi le plus souvent. Entre la justification et l'adhésion vient s'intercaler la représentation que le sujet a de son propre stock doxastique. Et celle-ci a, à son tour, un lien assez fort avec la représentation qu'il a de lui-même. Le rôle de ce que j'ai appelé plus haut la dimension épistémique de la subjectivité peut être souligné si l'on insiste sur le profil modal de la conviction.

Lorsqu'on considère le devenir d'une croyance dans l'espace logique, on doit en effet distinguer un changement dans sa valeur de vérité – je crois qu'il pleut et il pleut, je crois qu'il pleut et il ne pleut pas – et un changement qui touche le sujet ou le porteur de la croyance. Mais, alors qu'il est aisé d'imaginer un individu possible croyant que p , il est plus délicat de donner un sens à la représentation, en première personne, d'une croyance différente de mes croyances actuelles que je tiendrais *possiblement* pour vraie. Un trait caractéristique de l'adhésion de l'esprit à un contenu dans le cas d'une croyance ou d'une conviction, c'est en effet son caractère actuel²². Une personne peut, bien entendu, se représenter elle-même et

22. Gil insiste lui-même sur cet aspect en dégageant, à propos de l'intuition, ce qu'il appelle la « présencialité ». Voir les pages 96 et suivantes de *La Conviction, op.cit.*

s'attribuer des croyances qu'elle n'a pas actuellement. Elle peut juger que cette attribution est possible ou ne l'est pas et, si elle l'est, la situer dans une région plus ou moins éloignée de l'espace logique. Mais le lien de ces représentations – rendues possibles par le recours à l'espace logique – avec la croyance proprement dite est loin d'être clair. On peut se demander si la distance au monde actuelle mesure toujours uniformément l'allégation et l'auto-attribution. On peut aussi s'interroger sur la portée actuelle de ces attributions conditionnelles.

Soit p un contenu susceptible de faire l'objet d'une adhésion. Et soit S un sujet qui envisage, en première personne, cette adhésion comme possible. Un principe très fort, accepté, semble-t-il par Gil, consiste à supposer qu'il n'y a pas de différence entre adhérer actuellement et adhérer possiblement, ou, au moins, que l'adhésion possible entraîne nécessairement l'adhésion actuelle. Avec « adhérer à p » et « tenir pour possible d'adhérer à p » s'entendant en première personne, on aurait :

Uniformité modale doxastique forte : si S tient pour possible d'adhérer à p , alors S adhère à p .

Cette uniformité modale doxastique a pour avantage qu'elle est cohérente avec le caractère actuel de l'adhésion. « Tenir pour possible d'adhérer » est une expression étrange. Une adhésion seulement possible a quelque chose d'inconsistant, si l'on entend par adhérer l'acte même d'allégation ou d'auto-attribution. L'adhésion possible est, si elle est quelque chose, un autre nom de l'adhésion. Il me semble néanmoins aisé de montrer que ce principe est trop fort. Lorsque Millikan, dans son expérience célèbre, attendait les résultats des observations du mouvement des gouttes d'huile à travers un champ électrique, il était certainement engagé profondément dans l'expérience, au point que l'espace logique était scindé pour lui en deux grandes régions, selon que les mesures obtenues seraient ou non les multiples d'un seul *quantum*. D'un côté – pour simplifier – les mondes possibles quantiques, de l'autre les mondes possibles non quantiques. Dans cette situation épistémique, Millikan ne *savait* pas que le monde actuel était un monde quantique, puisque c'est ce qu'il entendait établir par l'expérience. Par suite il tenait pour possible d'adhérer au contenu [le monde est non quantique]. Néanmoins il reste vrai qu'il n'adhérait pas à ce contenu, sans quoi il n'aurait pas conçu une expérience tout entière destinée à établir que les phénomènes électriques observables sont quantiques. D'une manière générale, lorsque nous sommes engagés dans une recherche ou dans une enquête, nous ménageons une adhésion intermédiaire ou seulement possible au moins pour les croyances qui font directement l'objet de notre enquête. On préférera donc un principe plus faible :

Uniformité modale doxastique faible : si S tient pour possible d'adhérer à p , alors S adhère à *possiblement* p .

Attendant les résultats de son expérience, Millikan tient pour possible d'adhérer à la croyance selon laquelle le monde est non quantique et adhère actuellement à la croyance selon laquelle le monde est *possiblement* non quantique. Le choix entre ces deux principes est plutôt une affaire de circonstance. Dans certaines situations épistémiques le sujet envisage le devenir de ses croyances à travers l'espace logique et forme de cette manière une méta-représentation qui, en retour, modifie son adhésion actuelle aux croyances de premier ordre. Dans d'autres cas au contraire, le voyage dans l'espace logique n'a pas beaucoup d'effet. Comprendre pourquoi il en est ainsi me semble une tâche philosophique importante.

Empruntons à Stephen Yablo la manière d'approcher les situations de ce genre en approfondissant les conditions de la concevabilité²³. Kripke, comme on sait, a montré que la concevabilité était faillible. Il arrive que quelque chose, disons E, semble possible alors que c'est un fait qu'un monde qui est E ou duquel E est vrai ne peut pas être. Les opposants à l'usage élargi de la concevabilité s'arrêtent ici. Un monde E semble possible et ne l'est pas. Donc du « semble possible » on ne doit rien attendre. Mais, selon Yablo, les amis de la concevabilité répondront qu'on n'a fait que la moitié du chemin, que l'échec de l'inférence du concevable au possible prend généralement une forme spécifique dont il y a encore beaucoup à dire. Ils proposent donc un principe (O) :

(O) si l'impossible E *semble* possible, il y a quelque chose, F, qui *est* possible et nous prenons la possibilité de F pour celle de E.

S'il y a quelque chose (E) que S croit et qui est une illusion dans les possibles – par exemple que l'eau aurait pu ne pas être H₂O – alors il y a un objet de croyance approchant et, lorsque la croyance est dirigée sur cet objet, elle n'est plus une illusion – par exemple (F) : que ce qui remplit les lacs et les océans sur terre aurait pu ne pas être H₂O. Le problème porte alors sur le lien de la croyance actuelle dans E et de la croyance possible dans F. Le sujet croit, à tort, que l'eau aurait pu être autre chose que H₂O (croyance actuelle) mais cette croyance fautive peut être aisément reliée à la croyance vraie (possible) que ce qui remplit les lacs et les océans sur terre aurait pu ne pas être de l'eau. La thérapie des illusions modale nous conduit par suite (i) à substituer un objet de pensée à un autre objet, différent du premier, mais qui n'est pas complètement disparate (ici les deux objets sont reliés par l'intension bidimensionnelle de l'intension d'« eau »), (ii) à envisager l'effet actuel de croyances seulement possibles c'est-à-dire exactement ce qui est en jeu dans le principe d'uniformité modale doxastique.

23. S. Yablo, « No Fool's Cold: Notes on Illusions of Possibility », *Two-Dimensional Semantics*, M. Garcial-Carpintero et J. Macia (dirs.), Oxford, Clarendon Press, 2006, p. 334.

Afin d'approfondir ce deuxième aspect, Yablo proposait d'imaginer un psychologue qui dirait à son patient quelque chose comme : « vous êtes convaincu que personne ne vous aime. Et moi je suis convaincu que c'est une illusion. Je vais vous présenter une explication du sentiment que vous avez – que personne ne vous aime – et cette explication ne va pas vous convaincre. Tout ce que je vous demande, c'est de vous représenter que cela pourrait être une explication du sentiment que vous avez ». Autrement dit, le psychologue propose au sujet de former la pensée :

Je suis *peut-être* convaincu que p ,

et espère que ce *peut-être convaincu* finira par avoir un effet sur ses croyances actuelles du sujet, en particulier sur sa croyance fautive selon laquelle personne ne l'aime. Yablo ne nous dit pas par quel chemin cognitif a lieu le miracle psychologique, s'il a lieu. Mais s'il a lieu, on a bien affaire à une adhésion, c'est-à-dire à un acte cognitif qui s'enracine dans la représentation d'un état mental seulement possible.

On peut se demander sur quelle dimension de l'adhésion le psychologue entend intervenir : sur l'allégation – il est possible que je tiens pour vrai que p – ou sur l'auto-attribution – c'est une version possible de moi-même qui tient pour vrai que p ? Je pense pour ma part qu'il y a là un dosage subtil, qui a lieu en nous sans nous, et dépend à la fois du contexte et du contenu envisagé. Quelques exemples suffiront pour finir.

Exemple 1. A est célibataire. Il croise une femme dans une soirée, elle renverse du vin sur sa chemise. Il est furieux, il lui dit de la manière la plus désagréable possible « vous pourriez faire attention » et il va nettoyer sa chemise comme il peut. Plus tard, A se représente la même soirée : la même femme aurait renversé le vin, il lui aurait souri pour la rassurer, elle lui aurait proposé de nettoyer la tache et il aurait dit oui, ils seraient allés tous les deux dans la cuisine et elle lui aurait raconté sa vie et ils seraient tombés amoureux l'un de l'autre. A se représente ce monde possible. Il se représente comme possible qu'il est amoureux de cette femme. Je veux bien croire que A n'est pas amoureux pour autant. Mais on accordera qu'il n'en est pas loin²⁴. L'amour ne voyage pas dans l'espace logique mais il y a quelque chose qui voyage à sa place : le regret, la nostalgie, la possibilité de l'amour.

Exemple 2. A est catholique. Néanmoins, A se représente comme possible un monde dans lequel il serait lui-même convaincu que Dieu ne s'est pas incarné. Est-ce que cela veut dire que A est actuellement convaincu que Dieu ne s'est pas incarné ? Évidemment non. Tout au plus peut-on conclure

24. Une célèbre parabole de Kafka, *Auf der Galerie*, explore cette dimension modale de l'amour : un jeune homme voit sur la piste du cirque une écuyère encouragée par un public nombreux et choisi dansant sur un cheval majestueux et sous le regard d'un Monsieur Loyal doux et paternel ; il imagine la même scène avec une pauvre écuyère sur un triste cheval tournant dans un cirque misérable et, de cette possibilité qu'il envisage, il s'ensuit que le jeune homme pleure.

que A se représente comme possible qu'il n'est pas catholique. Il y a un monde accessible à A dans lequel A n'est pas catholique. Mais la croyance ici ne voyage pas davantage dans l'espace logique. Elle n'a pas d'effet en retour. Rien de ce qui fait l'adhésion actuelle de A – si ce n'est peut-être une forme de tolérance dans sa relation aux autres²⁵ – ne se trouve modifié par la représentation d'un état différent de son stock de croyances fondamentales.

Exemple 3 (l'expérience de Nozick)²⁶. A mange de la viande. Il est convaincu qu'il a le droit de manger des animaux parce qu'il est beaucoup plus intelligent qu'eux, et parce que c'est nécessaire à son équilibre physiologique. Il s'imagine qu'il y a une sorte d'échelle des êtres et que, si l'on est vraiment plus haut qu'un être sur cette sorte d'échelle, on a le droit de le manger. Un jour A imagine qu'il rencontre des extraterrestres beaucoup plus intelligents que lui et beaucoup plus puissants, et qui ont la même notion de l'échelle des êtres. Celle-ci leur donne le droit de manger A et, comme il se trouve que A semble avoir bon goût, ils entreprennent de le manger. Placé dans ces conditions, A déclarerait qu'il ne veut pas être mangé et que cette histoire d'échelle des êtres est fallacieuse, que certes il y a une échelle des êtres mais que l'existence de cette échelle ne justifie nullement que certains mangent et que d'autres soient mangés. Ayant imaginé cette situation, A choisira-t-il d'être cohérent avec sa croyance seulement possible ?

Chacun de ces exemples mériterait qu'on s'y attarde. En particulier, il vaut la peine de chercher à caractériser une inertie de la représentation du sujet épistémique, inertie telle que certaines de nos croyances possibles n'ont aucun effet en retour sur l'adhésion actuelle. Quant à l'expérience de Nozick, elle est riche d'enseignements sur l'adhésion dans son lien à la rationalité. Il est en effet probable que, à la suite de cette expérience, A continue à manger de la viande, et que ses habitudes alimentaires ne soient que très superficiellement touchées par des arguments de ce genre. En revanche, la justification qui était celle de A avant l'expérience – il existe une échelle des êtres qui nous autorise à manger les animaux que nous mangeons – ne devrait pas lui survivre. A prendrait en effet conscience que c'est sa position sur l'échelle actuelle des êtres, qui fait de lui un individu très favorisé et qui détermine actuellement son adhésion à la croyance selon laquelle l'échelle en question fournit une justification satisfaisante à sa pratique alimentaire. Dans la vie rationnelle, nous cherchons à maximiser autant que possible l'uniformité modale doxastique.

Jean-Baptiste RAUZY
Sorbonne Université

25. Je n'aborde pas ce point qui concerne plutôt la question du pluralisme.

26. Voir R. Nozick, *Anarchy, State and Utopia*, New York, Basic Books, 1974, pp. 39 sqq.

