



Chapitre de livre

1998

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

---

## Dispositions à agir et volonté de croire

---

Engel, Pascal

### How to cite

ENGEL, Pascal. Dispositions à agir et volonté de croire. In: Subjectivité et conscience d'agir: approches cognitive et clinique de la psychose. J. Proust et H. Grivois, (Ed.). Paris : PUF, 1998. p. 115–138. (Psychologie et sciences de la pensée)

This publication URL: <https://archive-ouverte.unige.ch/unige:5022>

## DISPOSITIONS A AGIR ET VOLONTE DE CROIRE\*

Pascal Engel

Université de Caen, CREA and Institut Universitaire de France

*Version antérieure d'un texte paru dans H. Grivois et J. Proust, ed. Subjectivité et conscience d'agir, PUF 1997*

Croire, avoir des croyances, est quelque chose qui paraît relever de la partie passive de notre esprit. Et pourtant il semble que nous puissions aussi exercer une forme de contrôle sur nos croyances: nous les affirmons, nous y adhérons, ou nous les rejetons pour en affirmer d'autres. Il y a des choses que nous aimerions croire, et sans doute aussi que nous aimerions ne pas croire. Nous pouvons aussi prendre nos désirs pour des réalités, et même, semble-t-il, nous mentir à nous-mêmes. Et pourtant peut-on vouloir croire, et réussir à croire quelque chose par l'effet de notre volonté, c'est-à-dire décider de croire? La méthode Coué a ses limites, et les conflits ressentis par le sujet qui a formé une croyance dissonnante, irrationnelle, ou pathologique— ou même simplement celui qui se définit comme un “croyant”— montrent qu'il n'est pas facile de croire quelque chose à “tout prix” ou contre l'évidence, simplement parce qu'on a désiré le croire.

Au risque de décevoir un lecteur qui souhaiterait que l'on aborde directement de telles questions en s'appuyant sur les données cliniques provenant, par exemple, des travaux sur les croyances délirantes ou la dissonance cognitive, je voudrais d'abord les examiner en me plaçant en quelque sorte en amont, en considérant les propriétés les plus élémentaires de ce que l'on appelle ordinairement des “croyances”. Est-il réellement possible de croire “à volonté”? Je voudrais d'abord montrer que la définition la plus plausible que l'on peut donner de la notion de croyance et de ses propriétés, selon laquelle les croyances sont des dispositions à agir, rend extrêmement problématique une telle possibilité. Mais je voudrais aussi, dans un second temps, soutenir qu'il y a un sens dans lequel nous pouvons parler d'un contrôle de la volonté sur nos croyances, bien qu'il ne s'agisse plus, dans ce cas, de

---

\* Une première version de cet article a été lue au colloque organisé par Joelle Proust et Henri Grivois sur “Subjectivité et conscience d'agir” en avril 1996. Je les remercie, ainsi que les participants à cette rencontre, pour leurs remarques. Une version plus ancienne avait été esquissée lors d'un séminaire de D. Widlöcher à la Salpêtrière en 1995 où j'ai parlé des croyances délirantes, et pour lequel je remercie M.C Hardy de ses remarques.

croyances au sens dispositionnel, ni d'une volonté qui aurait un contrôle direct sur nos actions. Cela ne réfute pas la conception dispositionnelle, mais doit nous conduire à l'élargir, en particulier de manière à ce qu'elle inclue ce que l'on pourrait appeler la *structure subjective* d'un agent, ses plans, ses objectifs, et ses buts. Ce n'est que dans un troisième temps que j'essaierai d'appliquer ces remarques à des phénomènes à la formation de croyances "anormales": la *self-deception* et les "croyances délirantes".

### ***1. La thèse dispositionnelle-fonctionnelle***

Dans son livre *The Emotions and the Will* (1859) le philosophe et psychologue écossais Alexander Bain <sup>1</sup> donna une définition de la croyance qui devint fameuse à travers l'élaboration qu'en proposa le fondateur du pragmatisme, C.S. Peirce <sup>2</sup>, et plus tard, durant les années 1930, à travers les œuvres des philosophes de Cambridge influencés par le pragmatisme, F. Ramsey et R. Braithwaite <sup>3</sup>. Selon cette définition, les croyances sont des *dispositions à agir* ou des habitudes d'action. Sous sa forme la plus élémentaire, la conception dispositionnelle des croyances compare les croyances à des dispositions physiques, comme la solubilité, la fragilité, la friabilité, et ainsi de suite. Ce genre de termes dispositionnels sont en général analysés en termes de conditions contrefactuelles du genre "X est D-ible = X D-raït si telles ou telles circonstances se réalisaient" (par exemple: "X est soluble = X se dissoudrait si X était plongé dans tel liquide). Comme on le sait, ce genre de définitions pose toutes sortes de problèmes, et dépend de la manière dont on comprend la notion de disposition et les conditions contrefactuelles qui sont supposées la définir.<sup>4</sup> Selon une certaine conception des dispositions, une propriété dispositionnelle d'une chose n'asserte rien d'autre que des propriétés d'un comportement observable, ou d'un comportement observable possible, et c'est pourquoi la théorie en question a souvent été mise au service d'une forme de behaviorisme. En ce sens nous pourrions "définir" la croyance de Jean qu'il pleut ainsi:

---

<sup>1</sup> Bain 1859. Sur Bain, voir les remarques de M. Jeannerod in Jeannerod 1996, p.94-96. Il remarque en particulier que Bain, à la différence des associationnistes classiques ou "sensualistes" comme Mill, insistait sur l'importance de la volonté sur l'activité mentale, et soutenait que "l'action est une propriété plus intime et plus inséparable de notre constitution qu'aucune autre de nos sensations" (cité par Jeannerod, p.96). C'est précisément pourquoi il est quelque peu ironique que la théorie de Bain puisse apparaître, comme je le montre dans la suite de cet article, comme un obstacle à l'idée que la volonté puisse influencer sur la formation des croyances.

<sup>2</sup> Sur le sens de cette thèse chez C.S. Peirce, cf. Engel 1983 et cf. Tiercelin 1993.

<sup>3</sup> Ramsey 1926 (1991), Braithwaite 1945. Voir Engel 1983, et à paraître, sur les idées de Ramsey.

<sup>4</sup> voir par exemple Goodman 1956. Sur l'analyse qui suit, cf. également Engel 1994, p.22-24

- (1) Jean croit qu'il pleut si et seulement si, si le comportement C était observable, Jean ferait X, si le comportement D était observable, Jean ferait Y, si le comportement E était observable, Jean ferait Z, etc.

Mais ce genre de définition se heurte à plusieurs difficultés. Tout d'abord le "etc." indique que la liste est potentiellement ouverte, car, comme Ryle (1949) le note, on ne peut pas limiter la liste de comportements caractéristiques à un seul: il n'y a pas de limite précise à ce qu'un individu est susceptible de faire quand il croit qu'il pleut, ce qui veut dire que les croyances ne sont pas des dispositions à agir d'une seule sorte de façon, mais d'un grand nombre de façons. Cela fait obstacle à une véritable définition behavioriste. Ensuite (1) présuppose que le *definiendum* ("croit qu'il pleut") est un terme mental, et que le *definiens* (la liste ouverte de comportements possibles) soient indépendants, c'est-à-dire qu'il n'y ait pas de termes mentaux dans cette liste. Mais supposons que C soit, par exemple, le comportement de prendre son parapluie. Or il semble que ce comportement ne soit une réponse appropriée à la croyance qu'il pleut que si l'agent *croit* qu'un parapluie protège de la pluie, et *désire* se protéger de la pluie, ce qui réintroduit dans le *definiens* non seulement des termes mentaux, mais également le terme même ("croit") qui devait être défini. La définition menace d'être circulaire. Enfin, on peut objecter que ce genre de définitions ne rend pas compte de notre intuition commune selon laquelle les croyances, les désirs, et d'autres états mentaux sont *causes* du comportement, et non pas simplement *associés à* ce comportement. Mais, sauf à admettre l'idée traditionnelle, aristotélicienne ou scolastique, selon laquelle les causes sont des "pouvoirs" ou des "facultés" dans les choses, une disposition n'est pas une cause: la solubilité du sucre n'est pas la cause du fait que le sucre se dissout; c'est la structure physique du structure qui est la cause, c'est-à-dire ce que les philosophes appellent la "base catégorielle" de la disposition <sup>5</sup>. Nous admettons certainement que

- (2) Si les circonstances C, D, E, etc. étaient présentes, Jean ferait X ( ex. prendrait son parapluie), X, Z, etc. *parce qu'il* croit qu'il pleut.

Mais si nous adoptions une définition de type (1), (2) devrait se traduire en:

- (3) Si les circonstances C, D, E, etc. étaient présentes, Jean ferait X ( ex. prendrait son parapluie), X, Z, etc. *parce que* si les circonstances C, D, E, etc. étaient présentes, Jean ferait X ( ex. prendrait son parapluie), X, Z, etc.

---

<sup>5</sup> Voir par exemple Amrstrong 1973.

Or (3) est parfaitement vide et trivial. Si nous voulons conserver le caractère causal des explications ordinaires de l'action <sup>6</sup>, il semble que nous devions ou bien rejeter la conception dispositionnelle, ou bien la modifier substantiellement de manière à ce qu'elle rende compte de ce caractère causal.

Il y a une manière de suivre la seconde voie. Elle consiste d'abord à admettre que les croyances ne sont pas simplement des dispositions au comportement, mais des dispositions au comportement *étant donné d'autres états mentaux*, tels que des désirs, des intentions, ou d'autres croyances. Limitons-nous aux désirs. On révisé alors la définition (1) ainsi:

(4) X croit que P si et seulement si X est disposé faire une action A de manière à satisfaire son désir que Q si P était vrai.

La notion de désir elle-même est susceptible de recevoir le même genre de définition:

(5) X désire que Q si et seulement si X est disposé à faire une action A conformément au fait qu'il croit que P.<sup>7</sup>

Par exemple: Jean croit qu'il pleut (P), si et seulement si il est disposé à l'action A (par exemple à prendre son parapluie) de manière à ce que A satisfasse son désir (par exemple) de rester sec (Q). Et Jean désire rester sec (Q) si et seulement si il est disposé à agir conformément au fait qu'il croit qu'il pleut (P) (par exemple en prenant son parapluie (A)). Les deux définitions (4) et (5) doivent être conjointes. Et ainsi chaque concept à définir survient dans le *definiens* de l'autre: le désir sert à définir la croyance, et la croyance sert à définir le désir. Certes l'analyse en question est, de ce fait, circulaire. Mais ce n'est pas un défaut, si nous rejetons le behaviorisme de (1): nous admettons que nous ne pouvons pas donner une analyse réductrice de la croyance en termes de désir, ni vice versa. C'est cette idée qu'incorpore la conception selon laquelle croyances et désirs sont des états *fonctionnels*: une croyance conduit à des

---

<sup>6</sup> Bien entendu, c'est un prix que certains philosophes ne sont pas prêts à payer, car ils ne tiennent pas ces explications comme causales. Mais ici, je suppose simplement cette analyse causale comme correcte, et que les arguments de Davidson (1980) le sont aussi. cf. Engel 1996.

<sup>7</sup> Le "si et seulement si" est sans doute ici trop fort, car d'autres désirs ou d'autres croyances peuvent intervenir, qui ne conduiraient pas aux actions en question (par exemple Jean peut désirer rester sec, croire que la pluie mouille, mais avoir aussi une croyance religieuse que la pluie est sacrée, et désirer être au contact de ce liquide sacré tombé du ciel; en ce cas, il se laissera tremper. Ces conditions *ceteris paribus*, comme on les appelle, doivent limiter l'effet de la définition, en sorte que les formules (4) et (5) devraient être préfixées de: "typiquement X croit que ...". L'ajout de cette condition a pour effet de faire une pétition de principe quant aux *autres* désirs et croyances de l'agent, alors même que ce dont il va être question ici, ce sont ces autres facteurs (en particulier motivationnels qui entrent dans la production de croyances. Je reviens sur ce point plus bas.

actions en fonction de désirs, et un désir conduit à des actions en fonction de croyances, et croyances et désirs sont fonctions les uns des autres. Mais cela ne suffit pas, si nous voulons aussi incorporer l'idée que les états en question sont causaux. Il faudra spécifier aussi que les états en question sont des *fonctions causales* des désirs aux actions, et des croyances aux actions. On parvient ainsi à l'idée fonctionnaliste familière, selon laquelle croyances et désirs sont des fonctions ou des rôles causaux. Cette conception pose toutes sortes de difficultés, et il y a de nombreuses manières de la formuler. Mais pour les besoins de la présente discussion, je l'admettrai sous cette forme, et je l'appellerai la conception *dispositionnelle-fonctionnaliste* des croyances.<sup>8</sup>

## 2. *La volonté de croire ou les croilontés*

Le problème que je voudrais poser à présent est le suivant. Supposons que nous puissions décider de croire quelque chose, ou vouloir croire quelque chose, tout comme nous pouvons décider de lever le bras ou de boire un verre d'eau. En ce cas, croire que P ne serait pas un état dispositionnel-fonctionnel causant l'action. Ce serait en soi-même une action, si une décision ou une volonté ont une action comme objet et comme effet. Mais cette thèse entre en conflit avec divers faits *conceptuels*.

Elle entre d'abord en conflit avec la thèse selon laquelle les croyances sont des dispositions à agir, car les actions sont, au moins selon une conception plausible, des *événements*, des choses qui se passent dans le monde (Davidson 1980). Et les dispositions ne sont pas des événements, ni donc, en ce sens, des actes. Ensuite, cela entre en conflit avec la conception causale des actions: car, à nouveau selon une analyse plausible, décider de croire quelque chose, c'est se causer soi-même à croire quelque chose, et si les croyances sont des causes des actions, cela entraîne que l'on puisse se causer soi-même à causer ses propres actions. Mais comme l'a noté Davidson (1980), agir, ce n'est pas se causer soi-même à être la cause de ses actions. C'est simplement accomplir ces actions. Enfin, l'idée que la croyance puisse être une action entre directement en conflit avec un trait conceptuel qui distingue la volonté, ou l'intention, et les états motivationnels en général, des états cognitifs, tels que les croyances. En général une croyance est causée par l'existence d'un état de choses du monde (par exemple la croyance qu'il pleut est causée par le fait qu'il pleut), et a, pour user de l'expression d'Anscombe (1958), reprise par Searle (1983) une "direction d'ajustement" (*direction of fit*) qui va "du monde à l'esprit", alors qu'en général

---

<sup>8</sup> Voir aussi Engel 1995.

les actions sont causées par des intentions (volontés, etc.), et ont une direction d'ajustement qui va "de l'esprit au monde" (nos désirs sont *réalisés* par des états du monde, qui ne sont pas des conditions de *vérité* de ces désirs, mais des conditions de leur *satisfaction*, alors que nos croyances ont comme conditions de vérité des états de choses du monde, auxquels elles s'adaptent ou non, tandis que c'est le monde qui s'adapte ou non à nos désirs) Mais la conception dispositionnelle-fonctionnelle tient les croyances comme causées par des entrées d'information de l'environnement, et comme causant en retour (moyennant des désirs) des actions. Si les croyances pouvaient être volontaires, alors les croyances comme facteurs causaux pourraient elles-mêmes être causées par des actions de "croire", qui ne seraient par conséquent pas des entrées d'information externes, mais des entrées internes. En fait ce ne seraient même pas des entrées, mais, si l'on peut dire, des entrées "créées" par le sujet lui-même. Mais cette condition paraît impossible, en raison même de l'asymétrie entre les croyances et les désirs. Elle voudrait dire que toutes nos croyances sont des formes de désirs, ou des manières de prendre nos désirs pour des réalités.

C'est pourquoi nombre de philosophes ont été amenés à douter, *a priori*, de la possibilité même de "croire à volonté", au moins quand il s'agit des croyances empiriques ordinaires. Dans un essai classique, Bernard Williams déclare:

"Une idée centrale, au sujet des croyances empiriques est celle d'une croyance que P à laquelle on parvient parce que les choses sont ainsi, qui implique une relation entre l'environnement perceptif d'un individu, ses perceptions et les croyances qui en résultent. Si un concept ne satisfait pas aux exigences de cette notion, à savoir que nous puissions comprendre l'idée qu'il vient à croire que P parce qu'il en est ainsi, et parce que ses organes perceptifs fonctionnent normalement, on n'aura pas affaire au concept de croyance empirique... Mais un état qui pourrait être produit à volonté ne satisferait pas à ces exigences, parce qu'il n'y aurait pas de connexion régulière entre l'environnement, les perceptions, et ce à quoi l'individu a abouti, qui sont des conditions nécessaires pour qu'il y ait croyance." (Williams 1973, p.137)

En d'autres termes, c'est une vérité conceptuelle, et non pas empirique, que (i) les interactions sensorielles avec le monde sont les sources canoniques de nos croyances au sujet du monde, et (ii) que dans de telles interactions la croyance que P puisse provenir d'une source sensorielle à partir d'un état de choses dans lequel P est le cas, autrement dit que la personne croit que P parce qu'il en est ainsi. Or l'analyse dispositionnelle-fonctionnaliste serait fautive si les croyances n'obéissaient pas à ces réquisits conceptuels. Car si les entrées qui causent les croyances n'étaient pas des entrées sensorielles, mais des entrées

“voulues”, le concept de croyance ne pourrait jouer aucun rôle. Cela semble être la raison principale pour laquelle je ne peux pas croire à volonté, par exemple, que mon pantalon est en feu. Je ne peux croire que cela que si je vois, ou perçois, d’une manière quelconque (y compris hallucinatoire), que mon pantalon est en feu. Si ma croyance n’était pas produite de cette manière, ce ne serait pas une croyance. On peut exprimer la même idée en disant, comme Williams, que “les croyances visent à être vraies”: si je pouvais acquérir une croyance à volonté, je pourrais l’acquérir qu’elle soit vraie ou non. Mais il serait difficile de penser que dans ce cas on aurait affaire à une *croyance*.

Mais, comme l’a noté Jonathan Bennett (1990), cet argument n’infirmes la thèse selon laquelle les croyances pourraient être acquises à volonté que si l’on comprend (i) d’une certaine façon. Il contredit certainement cette idée au sens où les croyances contingentes portant sur des faits empiriques doivent fondamentalement être acquises à partir d’entrées sensorielles produites par l’environnement. Mais elle ne menace pas la possibilité qui est ici en question, qui est la possibilité d’induire des croyances *immédiatement* (“juste comme ça”). Car les croyances ainsi immédiatement induites pourraient être atypiques. Et ces croyances atypiques pourraient être des causes atypiques du comportement. Supposons donc que ce genre de “croyances” atypiques puissent causer des actions. Appelons des croyances qu’on pourrait acquérir volontairement des *croilontés*. Il semble y avoir deux manières d’acquérir de telles croyances.

L’une serait, comme on vient de le voir, une acquisition immédiate, comme si on pouvait acquérir une croyance en faisant claquer ses doigts. Cela placerait la croyance sur le même plan qu’une intention. Et il est clair que si une croyance est une disposition à l’action ou une habitude, c’est impossible: car on ne peut pas acquérir une habitude à la même vitesse que quand on claque les doigts. Mais cela n’exclut pas qu’une croilonté puisse être acquise *médiatement*, par l’intermédiaire d’autres événements, et au bout d’un certain temps. On ne peut acquérir une disposition à volonté “juste comme ça”, mais il semble que l’on puisse l’acquérir à la suite de la répétition d’un certain nombre d’événements, pour ainsi dire en la programmant.

Cette seconde interprétation est plus plausible en raison de la nature des dispositions. Comme on l’a vu, la thèse phénoméniste selon laquelle les dispositions sont seulement reliées à un ensemble de traits observables est douteuse, car une disposition doit être liée à une “base catégorielle”, dont elle dépend, au sens suivant: il ne peut pas y avoir de changement dans la disposition sans changement dans sa base catégorielle. Cela suggère que l’on ne

peut créer, ou changer une disposition déjà acquise qu'en modifiant la base catégorielle dont elle dépend, en rendant, par exemple, quelque chose fragile en altérant sa structure cristalline, ou en rendant quelque chose mangeable en le rechauffant, etc. En ce sens, on pourrait créer, ou altérer, en soi-même une croyance en altérant l'état catégoriel dont elle dépend. Comme on ne peut pas immédiatement le faire, les croyances induites immédiatement semblent impossibles, mais des croyances induites médiatement à travers les changements de leurs bases catégorielles semblent possibles.

Dans un article récent Michael Losonsky (1995) a proposé d'utiliser la distinction de Dretske (1988) entre des *causes structurantes* et des causes *déclenchantes* pour autoriser la thèse selon laquelle les croyances pourraient être induites médiatement. Selon Dretske, il y a au moins trois candidats possibles pour l'analyse d'un acte particulier, comme lever son bras. Il faut d'abord distinguer deux choses:

- (a) le mouvement du lever de votre bras.
- (b) le mouvement du lever de votre bras produit par une cause interne

Dans ces deux cas l'acte est identifié à un événement — le mouvement de votre bras. Dans (b), c'est un type particulier d'événement, celui qui a la cause appropriée (à la différence, par exemple, du lever de votre bras causé par une cause extérieure, par exemple quand quelqu'un saisit votre bras et le lève). Si l'on comprend les croyances en ce sens, alors, comme on l'a déjà fait remarquer, il est difficile de considérer les croyances comme de simples événements de type (a) ou (b). Mais Dretske défend l'existence d'une troisième forme de comportement intentionnel:

- (c) la production du mouvement par une cause interne

où la production en question n'est pas un événement simple, ni même un événement tout court, mais un *processus*, composé d'une *structure* d'événements, dont le mouvement de votre bras et sa cause interne font partie. Par exemple le fait de tourner l'interrupteur cause l'événement de l'éclairage de la pièce, mais c'est l'installation du dispositif électrique par l'électricien qui permet ce genre d'événement. La première cause est "déclenchante" et la seconde est "structurante". Appliquée aux croyances, cette distinction suggère que les croyances pourraient être induites si l'on induit des structures, et en ce sens, il semble que l'on puisse volontairement décider d'acquiescer une habitude de pensée, sans pour autant agir directement sur elle. Cette habitude pourrait

être acquise par un changement dans la base catégorielle, comme quand on prend des drogues, mais on pourrait supposer que l'hypnose, l'autosuggestion, l'incantation d'une phrase, ou d'autres actions, peuvent produire les mêmes effets. Quand Pascal, dans la seconde partie du célèbre argument du pari, suggère que ceux qui ne sont pas convaincus par l'argument pourraient agir comme des croyants, aller à la messe, prendre de l'eau bénite, et que "cela [les] abêtira", il suggère quelque chose de ce genre.

Mais à la réflexion, il n'est pas du tout évident que des croilontés "médiates" soient réellement des croyances motivées ou volontaires. Bennett (*ibidem*) fait une distinction utile, entre l'*immédiateté motivationnelle* d'une croyance, et son *immédiateté ontologique*. Le lever de votre bras, par exemple, en réponse à une récompense suffisamment motivée (par exemple si je vous paie 5000F pour le faire) est volontaire, bien qu'il s'effectue par la médiation d'une chaîne complexe d'événements neurologiques. Le geste accompli n'est pas ontologiquement immédiat, bien qu'il soit motivationnellement immédiat. Quand je vous offre une forte récompense pour lever votre bras, et que vous le levez "juste comme ça", vous ne vous provoquez pas à faire quelque chose qui aura comme résultat le lever de votre bras, et en ce sens mon offre de récompense ne se lie pas motivationnellement immédiatement au lever de votre bras, dit Bennett. Aucune pensée motivante au sujet des événements neurologiques qui sont la médiation ontologique de votre acte n'intervient dans le syllogisme pratique qui vous conduit de la prémisse selon laquelle vous acceptez la récompense à l'action qui en est la conclusion.<sup>9</sup>

C'est pourquoi la thèse de Losonsky selon laquelle on pourrait dire que les croyances peuvent être volontaires médiatement comme causes "structurantes" est ambiguë. Selon lui, le désir et la volition peuvent déterminer une structure de croyance tout comme un ingénieur peut déterminer la structure d'un thermostat. Mais que veut dire ici "structure"? Si par "structure de croyance" on entend la base catégorielle de la disposition à croire, alors nous n'avons aucune prise motivationnelle sur cette structure. Ce n'est que la structure causale, ou ontologique, de la disposition qui est en jeu. Et sur celle-ci, par définition, nous n'avons aucun contrôle. Par conséquent cela ne conforte en rien l'idée selon laquelle les croyances pourraient être acquises à volonté. Pour pouvoir avoir un contrôle volontaire, même médiat, sur nos croyances, il faudrait que nous ayons un contrôle sur leurs causes. Même si l'on suppose, par exemple, que le fait de prendre une drogue pourrait nous induire à acquérir certaines croyances, par exemple la croyance que nous pouvons voler dans les

---

<sup>9</sup> Bennett 1990, *ibid.*

airs, on n'aura une prise motivationnelle sur les *effets* de la drogue, mais pas sur ses *causes*. Et ce n'est que cette dernière qui pourrait nous permettre de dire que la croyance est volontaire, une croilonté authentique. Et la structure causale en question, même si l'on avait accès à elle, resterait ce qu'en dit Williams, c'est-à-dire qu'elle aurait la direction d'ajustement "du monde à l'esprit", et non pas "de l'esprit au monde".

### **3. La structure subjective et la politique du croire**

Il y a cependant encore un sens dans lequel on pourrait parler d'un contrôle volontaire sur nos croyances: quand la "structure de croyance" n'est pas causale, mais rationnelle, c'est-à-dire quand elle porte sur les relations qui existent entre des croyances non pas comme des états dispositionnels ou fonctionnels, mais comme *contenus propositionnels*, objets des croyances. Il ne s'agit plus alors des croyances comme *causes* d'actions, mais des croyances comme *raisons* d'avoir d'autres croyances. Si j'ai une raison pour agir ou pour croire, il semble bien que je puisse avoir une prise sur ma croyance, au sens où j'ai de bonnes raisons de l'affirmer, et, sur la base de ces raisons, de la défendre ou de la rejeter. Et cela semble pouvoir être l'effet d'un choix ou d'une décision volontaire. Mais il importe, ici encore, d'être clair sur la manière dont la "volonté" peut ici intervenir.

Car même si l'on croit pour certaines raisons, les raisons des croyances sont habituellement constituées par les *données* dont on dispose, qu'elles soient empiriques ou non. Pour le dire comme Williams, c'est une vérité conceptuelle que nos croyances sont façonnées par nos données. C'est, là encore, une simple conséquence du fait que les croyances aient la direction d'ajustement du monde à l'esprit, et proviennent normalement de sources sensorielles causées elles-mêmes par l'environnement. Le fait que nous ayons des raisons de croire, et que nous soyons, de ce point de vue, libres, comme le disait Descartes, d'affirmer ou de nier ce que nous venons à croire, introduit bien un élément volontaire, mais cet élément n'intervient que dans les *jugements* que nous formulons à partir de nos croyances: il n'intervient pas dans la source des croyances elles-mêmes, qui est en est habituellement la cause, et il ne change rien au fait que les raisons sont habituellement façonnées par les données disponibles. Il est certainement possible de distinguer en ce sens nos croyances, qui sont des causes d'action et qui ont des sources causales, d'autres types d'états mentaux que nous pouvons appeler, avec les philosophes classiques, des jugements, ou, pour utiliser une terminologie due au philosophe Jonathan Cohen, des

*acceptations*.<sup>10</sup> Je n'élaborerai pas ici la distinction de Cohen, mais on peut la motiver intuitivement. C'est une chose que, par exemple, de croire que Chirac est un homme sympathique, et c'est en une autre que d'accepter que Chirac est un homme sympathique. Croire cela, c'est, comme on l'a vu, être disposé à agir — par exemple à voter, conformément à cette croyance. L'accepter, c'est aussi être prêt à défendre cette croyance, à la prendre comme prémisse dans des raisonnements, et en général, adopter ce que l'on peut appeler une *politique épistémique* vis à vis de cette croyance<sup>11</sup>. Plus intéressant encore, on peut accepter, pour les besoins d'une hypothèse ou d'un raisonnement par l'absurde, une proposition P, sans pour autant croire qu'elle est vraie. Accepter que P, c'est au moins se comporter *comme si* on croyait que P, et être prêt à maintenir que P, ou à le prendre pour acquis, même quand les données ne sont pas favorables, ou sont contraires. En ce sens je peux accepter que Chirac est un homme sympathique sans le croire vraiment, et même agir en conformité avec cette acceptation (par exemple en votant pour Chirac), mais même si le résultat de mon action peut être le même que celui auquel me conduit la croyance correspondante, ce n'est pas la même action: car dans un cas (celui de la croyance) j'ai voté pour Chirac par conviction (ou par naïveté crédule), dans l'autre j'ai voté pour lui, mais sans me faire beaucoup d'illusions, peut être même cyniquement. Le point important, au sujet des acceptations, est qu'elles ne sont pas simplement tournées vers les données passées qui les justifient, ou les données présentes, mais vers les données futures. Accepter que P, c'est s'engager, dans le futur, à croire que P, et donc projeter la vérité et la justification de P. Le type de structures doxastiques auxquelles on a affaire avec les acceptations est apparemment fort distinct du type de structure propre aux croyances *simpliciter*. Je dirai que ce n'est pas une structure causale, mais une structure rationnelle, au sens où l'agent qui accepte certaines propositions se tient lui-même comme un agent rationnel, et comme responsable de ses acceptations, en un sens où un sujet qui croit n'a pas à le faire. Et si être un sujet, c'est se tenir comme responsable de ce que l'on croit, agent d'une certaine politique épistémique, alors la structure des acceptations recouvre une certaine *structure subjective*. Ce n'est pas incompatible avec la thèse dispositionnelle-fonctionnaliste, car nous pouvons dire que les croyances qui font l'objet d'un assentiment réflexif de la part d'un sujet sont ses dispositions à agir, *en tant qu'agent rationnel*. Ici, en quelque sorte c'est l'agent qui se forge ses dispositions, par un choix rationnel.

<sup>10</sup> Cohen 1992. J'ai exposé la distinction de Cohen notamment dans Engel 1995, 1996 a, et 1995 a, et à paraître.

<sup>11</sup> On trouvera une analyse détaillée de ces "politiques" de croyance dans Helm 1995

Si c'est le cas, alors il semble qu'il y ait bien, en ce sens, un type d'attitudes doxastiques qui soit plus ou moins sous le contrôle de la volonté, dans la mesure où accepter une certaine proposition est plus une sorte d'acte ou de décision qu'une forme de cognition structurée causalement. Mais s'agira-t-il vraiment de cas de croyances volontaires ou de croilontés au sens que nous envisagions ci-dessus? Toute la difficulté réside dans la question de savoir si des croyances-acceptations, ou les croyances subjectives en ce sens, peuvent réellement soustraites au contrôle de l'évidence. Et c'est ce dont il y a des raisons de douter, si l'on examine une version récente d'un argument destiné à établir qu'on peut vouloir croire, dû à Bas Van Fraassen (1984).

Van Fraassen part de la conception canonique de la manière dont les croyances sont reliées aux données qui les confirment qu'on trouve dans la théorie "bayésienne" de la confirmation <sup>12</sup>, qui n'est, en un sens, qu'une variante de la conception dispositionnelle-fonctionnelle. Selon cette conception, les croyances d'un agent sont ce sur la base de quoi il est disposé à agir, étant donné ses désirs et ses utilités. L'agent agit, en fonction de ses croyances et de ses désirs, d'après la règle selon laquelle il doit maximiser son utilité subjective ou espérée. Les croyances, en ce sens, correspondent à ses estimations de probabilité quant à l'occurrence de certains événements, et les désirs correspondent à ses degrés mesurables d'utilité. A chaque croyance correspond un degré de probabilité "subjective", qu'on peut établir, si l'on connaît le comportement de l'agent face à des paris. Le bayésien postule aussi que si l'agent est rationnel, il devra satisfaire certaines contraintes minimales de cohérence quant à la structure de ces assignations de probabilités et d'utilité (sans quoi, nous disent les bayésiens, on pourra faire contre lui des paris qui auront pour effet de le faire perdre systématiquement).

Admettons ce cadre général, et supposons à présent qu'un agent n'a pas que des croyances dues à ses interactions passées et présentes avec son environnement, mais aussi des croyances portant sur ses croyances *futures*. Un agent capable de projeter l'état de ses croyances passées et présentes sur l'état futur de ces croyances, et d'avoir une idée de leur cohérence à plus ou moins long terme, est ce que nous avons appelé un agent qui a une politique épistémique, un sujet. Il doit non seulement se projeter dans le futur, mais aussi avoir une certaine idée de la manière dont ces projections pourraient rétroagir sur ses croyances présentes ou passées, et le conduire à les réviser. A présent, supposons qu'un agent croie aujourd'hui, en  $t$ , à un degré de probabilité (ou de

---

<sup>12</sup> dont Ramsey est l'un des (re-) fondateurs contemporains. Voir les références citées dans Engel, à paraître a.

confiance)  $x$ , que P. Supposons qu'il découvre que, dans un temps futur  $t'$ , disons dans un an, il croira que P à un degré  $y$  *distinct* de  $x$ . Van Fraassen soutient que si l'agent est rationnel, son degré de confiance en la vérité de P en  $t$  ne devra pas différer, ou seulement très faiblement, de son degré de confiance en P aujourd'hui. En d'autres termes, le degré de confiance de l'agent en P en  $t$  ou la probabilité subjective présente de P, sous la supposition que sa confiance en P en  $t$  sera égale à  $y$  en  $t'$ , devrait être égal à  $y$ . C'est ce que Van Fraassen appelle le "Principe de réflexion".

Pour l'illustrer, supposons par exemple qu'un météorologue annonce aujourd'hui qu'il y a 50% de chances pour qu'il pleuve demain après midi, mais affirme aussi qu'il prédira demain matin qu'il y a 25% de chances pour qu'il pleuve demain après-midi. Un tel agent violerait le principe de réflexion, et apparaîtrait certainement aussi irrationnel, ou du moins étrange, que quelqu'un qui dirait: "Il pleuvra, mais je ne crois pas qu'il pleuvra"<sup>13</sup>. Le fait qu'il ait quantifié son degré de croyance ne le placerait sans doute pas face une forme de contradiction de ce type, et les degrés de croyance peuvent être vagues, mais même si sa prédiction est vague, il y aura au moins une tension caractéristique entre ses affirmations.

Le principe de réflexion semble être une maxime appropriée pour un agent rationnel capable de considérer ses croyances futures, et de projeter dans le futur ses degrés présents de croyance. Il semble indiquer une contrainte épistémique forte pour le sujet. Ici l'agent doit *vouloir*, s'il est rationnel, que son degré de croyance présent s'accorde avec son degré de croyance futur. Et pourtant le principe de réflexion semble pouvoir, et même devoir, être violé dans certains cas. Supposons — et c'est ici que nous retrouvons la question déjà abordée plus haut — que l'on sache que le fait de prendre une certaine drogue aura comme effet de me faire croire que je peux voler. Il serait certes irrationnel pour moi de croire *maintenant* que, puisque je croirai que je peux voler quand j'aurai pris la drogue, je peux voler. Il est au contraire rationnel, dans ce cas, de *résister* à ses croyances futures, et donc de croire qu'on ne volera pas (car toutes les données confirment le fait que nous ne sommes pas des êtres volants). De même Ulysse sait que le fait d'entendre le chant des sirènes le poussera à désirer les rejoindre dans une mer dangereuse, et lui sera fatal<sup>14</sup>. C'est pourquoi il s'attache au mât, et met de la cire dans les oreilles de ses marins, pour résister à son désir futur. Dans le cas qui nous occupe, il s'agit de résister à une croyance future. Il ne s'agit pas de rationalité pratique, mais de rationalité

<sup>13</sup> Ce genre de phrases est connu sous le nom de "paradoxe de Moore". Van Fraassen, dans d'autres écrits, rapproche son cas systématiquement de ce genre de paradoxe, mais je n'examine pas ce point ici.

<sup>14</sup> C'est J. Elster qui a attiré l'attention sur ce problème dans Elster 1982.

épistémique; mais puisque nous parlons de politique épistémique, les deux doivent obéir à des principes symétriques. L'avis "volontariste" que Van Fraassen nous propose pour nos croyances n'est qu'une contrepartie du problème pratique au sujet des désirs auquel est confronté Ulysse. Le principe de réflexion nous dit que nous ne devons pas faire confiance à nos opinions présentes si nous savons qu'elles entreront en conflit avec nos opinions futures. Mais au contraire, des cas comme celui de la drogue montrent qu'il faut faire confiance à nos opinions présentes, et non pas à nos opinions futures.

On répondra sans doute que l'exemple de la drogue est justement un cas déviant, où nous n'avons pas de contrôle sur nos croyances, parce qu'elles sont soumises à des causes sur lesquelles nous n'avons pas prise. Un sujet dont le système cognitif fonctionne bien, et qui a la structure réflexive appropriée, ne devrait pas déroger au principe de réflexion. Mais cette réponse se heurte exactement au même problème que celui que nous avons soulevé au sujet des croyances "volontaires" induites médiatement. Car s'il semble simple de distinguer le cas du sujet normal de celui d'un sujet soumis à des forces causales anormales, comme des drogues, est-ce toujours possible en principe? Comment pouvons-nous savoir si nos croyances futures ne seront pas soumises à des forces causales sur lesquelles nous n'avons pas le contrôle, et comment faire le partage, dans nos opinions, entre ce qui est causé par des chaînes causales normales et des chaînes causales déviantes?<sup>15</sup> Les causes sont motivationnellement médiates. Le principe de réflexion est justifié sur un plan idéal ou normatif, mais il suppose que nous avons le moyen de faire ce partage, et ce partage est lui-même une condition normative de la rationalité. On doit donc conclure sur ce point que ce principe ne valide pas l'idée d'une forme de croyances volontaires provenant de la structure subjective d'un agent rationnel, pas plus que la possibilité de "croilontés" médiates ne validait cette possibilité. Nos croyances, même rationnelles, restent sous l'effet de structures causales sur lesquelles un agent ne peut pas avoir totalement prise, même s'il exerce sur elle sa volonté.

#### ***4. Self-deception et croyances délirantes***

Ce n'est qu'à présent, après ce long détour, que nous pouvons essayer de tirer quelques leçons de cette discussion pour le problème des croyances formées de manière anormale, soit sous l'effet d'une *self-deception*, soit sous

---

<sup>15</sup> Notons, ici encore, le parallélisme entre le problème des chaînes déviantes cognitives et le même problème en philosophie de l'action. Voir Davidson 1980 et Searle 1983.

l'effet de quelque structure psychotique. Ces deux types de croyances sont ordinairement définies, aussi bien dans la littérature psychopathologique que philosophique sur ces sujets, comme des croyances que l'agent forme "en dépit des données disponibles", ou qu'il est prêt à croire "à tout prix" quelles que soient les données contraires.<sup>16</sup> Et quand on cherche à les expliquer, deux sortes d'hypothèses viennent naturellement à l'esprit. La première consiste à dire que l'agent est ici sous l'effet de quelque désordre cognitif, qui affecte la manière dont il reçoit les données, en sorte qu'il n'est pas en mesure d'ajuster correctement les jugements qu'il porte avec les sources sensorielles qui les produisent. Selon cette hypothèse, c'est la chaîne causale qui va de l'environnement aux récepteurs sensoriels et aux autres processus de traitement de l'information distale et proximale qui dévie quelque part. C'est donc d'abord un désordre dans la structure *subpersonnelle* ou *subintentionnelle* de l'agent qui expliquerait ces croyances. Rien, bien entendu, dans une telle hypothèse, ne peut conforter l'idée que l'agent influencerait sur ces croyances par l'effet de quelque état motivationnel. Mais une seconde hypothèse n'est pas moins naturelle. Elle consiste à soutenir que ces croyances où l'on s'illusionne, ou se trompe *soi-même*, sont bien des croyances qui sont l'effet d'un état motivationnel du sujet, qui, en quelque sorte, décide de croire malgré les données contraires dont il dispose. Si c'est bien ainsi qu'il faut les décrire, ces croyances seraient des contre-exemples nets à la thèse selon laquelle on ne peut croire "à volonté". Et c'est bien pour telles qu'on les tient, quand on examine cette possibilité.

Selon la définition de la *self-deception* qui est devenue plus ou moins canonique, celle de Davidson (1985), c'est ce qui se passe quand un agent se trompe ou se dupe lui-même. L'agent (par exemple un mari jaloux) pour certaines bonnes raisons, croit que P (par exemple que sa femme le trompe). Mais il désire que non-P (que sa femme ne le trompe pas). On suppose alors que son désir que non-P le conduit à construire la croyance que non-P, et à y adhérer. Mais un simple désir que non-P ne suffit pas à expliquer la formation de la croyance irrationnelle en question. L'agent doit aussi transformer son désir que non-P en croyance que non-P. Comment? Le modèle de Davidson semble reposer sur l'idée qu'en outre l'agent doit accomplir un certain acte, qui le conduit à croire que non-P, c'est-à-dire à *avoir l'intention* de faire quelque chose, et réussir à accomplir ce qu'il a l'intention de faire<sup>17</sup>. Quoi? Croire que

---

<sup>16</sup> cf. Maher 1974, Oltmanns & Maher 1988, Hemsley et Garety 1996, Proust 1995, et divers essais dans Proust ed.1996..

<sup>17</sup> Cette lecture de la position de Davidson ne m'était pas apparue clairement dans mes précédentes tentatives pour discuter le phénomène de la *self-deception*. (Engel 1995 b par exemple, où j'endorsse, sous P.Engel *Dispositions à agir et volonté de croire*

non-P, ou accepter que non-P. L'effet est qu'il croit que P (en vertu de ses données), mais aussi qu'il croit que non-P (contre ses données), d'où le conflit qu'il éprouve. Ce modèle semble donc bien supposer que quelque chose comme un vouloir-croire, une croilonté, est possible, et est responsable de la duperie ou de l'illusion en question. L'intention n'est pas nécessairement formée immédiatement; elle peut être formée médiatement, comme les croyances "pascalienues" que nous avons envisagées ci-dessus.

Si l'on suppose, par conséquent, qu'un agent peut réussir à se tromper ou à se duper lui-même (et non pas simplement essayer de le faire, et n'y parvient pas, ce qui est une autre lecture possible de ce qui se passe), il semble que le modèle usuel de la *self-deception* en fasse en désordre ou un conflit qui relève d'un échec de la rationalité pratique plutôt que de la rationalité cognitive, ou plutôt qu'il s'agit d'une certaine sorte, pour reprendre le terme employé ci-dessus, de politique épistémique suivie par l'agent, mais une politique épistémique foncièrement irrationnelle. Mais si les conclusions du paragraphe précédent sont correctes, cette analyse est fautive quand elle suppose que certains désirs de croire pourraient, par l'effet de certaines intentions, se transformer en *actions* de croire; or si l'on ne peut pas, même au sens d'une politique épistémique, authentiquement vouloir croire, il n'y a pas de raison pour laquelle la *self-deception* pourrait illustrer cette possibilité, car même les croyances déviantes obtenues de cette manière nous sont apparues impossibles. Il y a donc, sur ce point, de bonnes raisons de douter que la caractérisation davidsonienne du phénomène soit la bonne.

On peut cependant penser que Davidson lui-même ne souscrit pas complètement à cette implication de son modèle. Car il nous dit souvent seulement que c'est un *désir* qui cause la croyance contradictoire de la précédente, en laissant entendre que le désir opère de manière causale, comme une cause "non rationnelle". En ce sens, il rejoint le premier type d'explication, à cette nuance (importante) près qu'il tient la cause du désordre en question comme motivationnelle ou affective, et comme non cognitive. Le désir de croire que non-P conduit l'agent à construire des données qui le conduisent en retour à croire que non-P. Et Davidson suppose que ce processus se forme dans une partie de l'esprit du sujet distincte de celle dans laquelle se forme sa croyance que P, donc qu'il est un esprit "divisé". Il rejoint ainsi, comme il le note lui-même, des explications de ce genre de phénomènes comparables à celles que donne Freud, à cette différence près que le désir qui est à l'œuvre, selon Freud,

---

la forme de la notion d'acceptation, une version de la thèse selon laquelle la self-deception se forme sous l'effet d'une intention d'agir). C'est une conférence (encore inédite) d'Ariela Lazar en 1996 à Karlovy Vary en l'honneur de Davidson qui a attiré mon attention sur ce point.

est inconscient, alors que pour Davidson il ne l'est pas nécessairement. Dans le cas des croyances délirantes formées "contre l'évidence", Freud invoque, pour les neuropsychoses de défense, le mécanisme de la "projection", qu'il met en parallèle avec celui du refoulement qu'il invoque dans le cas de la névrose obsessionnelle (Widlöcher 1994, p.262 sq.<sup>18</sup>).

Même si cette seconde interprétation, ainsi que l'interprétation freudienne dont elle se rapproche, insistent cette fois sur la nécessité de fournir une étiologie aux croyances formées dans des cas de *self-deception*, et par conséquent tempèrent fortement l'idée que de telles croyances pourraient se trouver sous le contrôle du sujet, ce qu'elles ont en commun est de supposer que ces sont des états motivationnels qui constituent cette étiologie. Mais si le phénomène de la *self-deception* est compris comme l'obtention d'une structure mentale qui est l'effet d'une sorte d'*action* du sujet sur lui-même, je crois avoir montré toutes les limites de ce genre de description. S'il est correct, comme je le pense, de dire que des croyances volontaires ne sont pas vraiment possibles, on ferait mieux d'envisager d'emblée d'autres types de descriptions. Car il reste, dans la description et l'explication des phénomènes concernés, d'autres explications possibles. L'une d'elle, moins explorée, consiste à suggérer que l'agent qui s'illusionne lui-même n'accomplit pas réellement une action, mais est sous l'effet d'un désordre *émotionnel*. Une émotion peut certes conduire à accomplir certaines actions, ou nous y disposer, mais elle peut survenir en l'absence d'actions. Et on peut dire que les émotions affectent la cognition sans qu'un agent ait des buts, des plans, ou des intentions.<sup>19</sup> Cette hypothèse conduit aussi à s'intéresser, beaucoup plus que ne le fait le modèle de Davidson, aux processus subintentionnels qui peuvent affecter les croyances d'un agent dans ce genre de cas, plutôt qu'à ces croyances elles-mêmes, définies sur le mode propositionnel ou intentionnel.

Les mêmes remarques pourraient s'appliquer aux croyances délirantes, qui ne sont pas le même phénomène que la *self-deception*, dans la mesure où

---

<sup>18</sup> D. Widlöcher décrit l'explication freudienne du mécanisme de la projection, par exemple d'une croyance paranoïaque (conçue selon les lignes de l'analyse du cas Schreber), comme l'effet d'une double transformation, à partir d'un désir homosexuel. "Première transformation: "Je l'aime" devient "Je le hais". Seconde transformation: "Je le hais" devient: "Il me haït". Mais comment s'effectuent les transformations en question? Comme le dit Widlöcher, la réponse de Freud n'est pas claire. Il propose de décomposer le processus en deux: "D'une part un fantasme inconscient qui inverse la proposition du fantasme premier, inversion rendue possible par le rôle du narcissisme et de l'ambivalence dans la formation du fantasme premier, et d'autre part une expression consciente, marquée par la subjectivation de proposition ( X est dans un rapport de destruction avec Y Ø Je le hais)." (p.265-66). Mais même ainsi, on ne comprend pas bien comment la projection peut produire la croyance délirante. Si c'est une action du sujet sur lui-même (ou d'une sous-partie inconsciente du sujet sur une autre consciente) on retombe dans les difficultés du modèle "pratique" de la formation de croyance délirante. Si c'est un processus purement causal ou mécanique, comment peut-il lier des "propositions"?

<sup>19</sup> Cette hypothèse est suggérée par Ariela Lazar dans la conférence citée ci-dessus.

un agent qui se trompe lui-même n'est pas nécessairement un agent qui délire, de manière pathologique, et réciproquement (un agent qui délire n'est pas nécessairement un agent victime de *self-deception* .<sup>20</sup> Mais ce que les deux ont en commun est que l'agent croit "contre l'évidence". Le DSM III définit en effet le délire ainsi:

"Croyance fautive d'un individu fondée sur une inférence incorrecte concernant la réalité extérieure, fermement entretenue en dépit de ce que presque tout le monde croit et en dépit de ce qui constitue des preuves ou des données incontestables et évidentes. La croyance n'est pas habituellement de celles qui sont acceptées par les autres membres de la sous-culture de l'individu".<sup>21</sup>

Mais on retrouve ici toutes les ambiguïtés que nous avons déjà relevées. Est-ce que "inférence incorrecte" renvoie à un facteur causal sur lequel l'agent n'a pas prise? Et si oui, quel est le rôle de la volonté de l'agent dans "l'entretien ferme" de cette croyance? Est-ce *lui* qui croit contre l'évidence ou une partie subintentionnelle de son système cognitif? Nos deux manières d'expliquer le phénomène reviennent. L'une tend à accentuer le rôle des désordres perceptifs et des erreurs de jugement à l'origine de ces croyances. Elle insiste sur le caractère cognitif anormal, et sur les facteurs de raisonnement irrationnel dans la formation et le maintien des délires, aussi bien du point de vue déductif qu'inductif. Elle en fait des désordres "froids", relevant de la partie cognitive de l'esprit, plutôt que "chauds", relevant des effets des affects et des motivations sur la cognition. L'autre insiste au contraire sur le caractère de certitude absolue et d'évidence (au sens usuel, cette fois, pas au sens anglais de "preuve") qu'éprouve le sujet de la croyance délirante. Elle tend à favoriser, du même coup, le rôle des facteurs motivationnels dans le maintien de cette croyance, comme relevant de facteurs subjectifs plus ou moins "centraux", et sur sa "subjectivation", en dépit éventuellement de ses sources subpersonnelles.

Je ne dis pas que les deux types d'explication s'excluent. Elles peuvent être compatibles, et c'est sans doute la tâche d'une théorie de la formation des croyances anormales que de chercher à intégrer ces deux dimensions. Mais je pense néanmoins que les limites conceptuelles que j'ai soulignées quant à la possibilité de la volonté de croire sont aussi des limites du modèle explicatif qui fait de ces croyances des effets de troubles exclusivement motivationnels. Le

<sup>20</sup> Il y a une ambiguïté dans le terme français. Comme le remarque Widlöcher (p.256, note), l'allemand a deux mots, *Wahn* et *Delirium*, comme l'anglais, qui a *Delusion* et *Organic Delirium*, alors que le français n'en a qu'un, *délire*. Cela peut expliquer que la littérature anglo-saxonne sur ces sujets ait tendance à regrouper sous la rubrique des *delusions* à la fois des phénomènes proprement psychotiques et des phénomènes moins marqués de croyances anormales, alors que l'emploi du mot "délirant" en français accentue d'emblée la première connotation.

<sup>21</sup> cité par exemple, par Proust 1995, p.116, Widlöcher 1994, p.250

fait que les croyances soient des dispositions à agir n'entraîne nullement qu'elles puissent être produites par des actions, des intentions, ou même des plans, conscients ou non, d'un agent, même si des plans peuvent se greffer sur elles à un niveau distinct de la vie psychique<sup>22</sup>.

J'ai soutenu que les conditions normales de formation d'une croyance, et les caractéristiques conceptuelles de cette notion posent des limites nettes à la possibilité d'un contrôle motivationnel de l'agent sur ses croyances, et qu'en ce sens une croyance, même incluse dans ce que j'ai appelé la structure subjective d'un agent, ne peut pas être volontaire, bien que la manière dont un agent est conduit à regrouper les données qu'il tire de ses croyances puisse avoir ce caractère volontaire et motivationnel: il se conçoit alors non pas simplement comme un récepteur de certaines informations, mais comme un sujet qui est le type de personne qu'il *devrait* être s'il venait à croire certaines choses, ou, pour le dire en d'autres termes, comme un "croyant". On pourrait me répondre que mon argument ne concerne que les croyances rationnelles, ou typiques, et qu'il n'empêche nullement que des croyances irrationnelles, formées anormalement ou autrement pathologiques ou atypiques, pourraient bien relever de facteurs motivationnels, et néanmoins toujours mériter le qualificatif de "croyances". En ce sens, on est souvent amené à dire que la "logique" du psychotique ou du délirant n'est pas la nôtre, et que c'est une erreur caractéristique d'un préjugé "rationaliste" ou cognitif" que d'essayer de l'enfermer dans le moule normatif de la rationalité <sup>23</sup>. J'ai voulu au contraire suggérer que même la formation de croyances irrationnelles ou déviantes est soumise à ces limitations rationnelles et cognitives, et que c'est peut-être la conscience, même très confuse, qu'a l'agent du fait que ces limitations existent qui est une source du conflit qu'il peut éprouver.

## Références

Anscombe, E. 1958 *Intention*, Blackwell, Oxford

---

<sup>22</sup> Hemsley et Gaterly (1996) , mais aussi, semble-t-il Widlöcher, soutiennent ainsi qu'une analyse des croyances délirantes doit être "multifactorielle" et incorporer à la fois une étiologie en termes de désordres "cognitifs" et une étiologie faisant place à "l'affect, l'auto-estime et la motivation" (p.113), tous facteurs qui relèvent de ce que j'ai appelé la structure subjective et personnelle, et non pas subpersonnelle de l'agent. Si je comprends bien, la théorie de la psychose de Grivois (1995) joue aussi sur ce double facteur.

<sup>23</sup> Pour d'autres raisons de résister à ce genre de thèmes, cf. Proust 1995.

- Armstrong, D. 1973 *Truth, Belief and Knowledge*, Cambridge, Cambridge University Press
- Bain, A. 1859 *The emotions and the will*, London, Parker
- Bennett, J. 1990 “Is Belief voluntary ?” *Analysis*, n° 100
- Braithwaite, R.B. 1933 “The Nature of Believing”, *Proceedings of the Aristotelian Society*, 33, 129-146, repr. in in A.P. Griffiths 1967, ed., *Knowledge and Belief*, Oxford, Oxford University Press.
- Cohen, L.J. 1992 *An Essay on Belief and Acceptance*, Oxford, Oxford University Press
- Davidson 1980 *Essays on Actions and Events*, Oxford, Oxford University Press, tr.fr *Actions et événements*, Paris, PUF 1993
- 1985 “Deception and Division”, in Elster, J. ed *The Multiple Self*, Cambridge, Cambridge University Press, tr.fr in Davidson, *Paradoxes de l'irrationalité*, L'Éclat, Combas 1991.
- Dretske, F. 1988 *Explaining Behaviour*, MIT Press, Cambridge Mass.
- Elster, J. 1982 *Ulysess and the Sirens*, Cambridge, Cambridge University Press, tr.fr. partielle, *Le laboureur et ses enfants*, Paris, Minuit 1991
- Engel, P. 1983 Croyances, dispositions et probabilités”, *Revue Philosophique*
- 1994 *Introduction à la philosophie de l'esprit*, Paris, La Découverte
- 1995 “Les croyances”, in D. Kambouchner, ed. *Les notions philosophiques*, Paris, Gallimard, vol.2.
- 1995 a “Croyance, jugement et *self deception*”, *L'inactuel*, vol.3
- 1996 *Philosophie et psychologie*, Paris, Gallimard
- 1996 a “Croyances collectives et acceptations collectives”, in R. Boudon, A. Bouvier, et F. Chazel, eds, *Cognition et sciences sociales*, Paris, PUF
- à paraître “Ramsey: croyance, vérité et probabilité”, in D. Vernant, ed. *Les années 30*, Paris, Vrin.
- à paraître b *Believing and Accepting*, actes du colloque de Caen 1995
- Goodman, N. 1956 *Fact, fiction and forecast.*, Bobbs Merrill, Indianapolis, tr.fr. *Faits, fictions et prédictions*, Paris, Minuit
- Grivois, H. 1995 *Le fou et le mouvement du monde*, Paris, Grasset
- Helm, P. 1995, *Belief Polices*, Cambridge, Cambridge University Press
- Hemsley, D.H. & Garety, P 1996. “Les croyances délirantes”, in Proust 1996
- Jeannerod, M. 1996 *De la physiologie mentale*, Paris, O. Jacob
- Losonsky, M. 1995 “Wanting to believe”, à paraître in Engel à paraître b
- Maher, B. 1974 “Delusional Thinking and Perceptual Disorder”, *Journal of Individual Psychology*, 30, 98-113, tr.fr. “Pensée délirante et désordre perceptif”, in Proust 1996
- Oltmanns, T. & Maher, B.A. 1988 *Delusional Beliefs*, New York, Viley
- Proust 1995 “Raisonnement, argumentation et rationalité chez le psychotique”, *Hermès*, 16, dir. A. Boyer, p.113-30
- ed. 1996 La folie dans la place, *Raisons pratiques*, n°7.
- Ryle, G. 1949, *The Concept of Mind*, Hutchinson, Londres, tr.fr. *Le concept d'esprit*, Paris, Payot 1976.
- Searle, J. 1983 *Intentionality*, Cambridge, Cambridge University Press, tr.fr. *L'intentionnalité*, Paris, Minuit 1987
- Tiercelin, C. 1993 *La pensée-signe*, Nîmes, J. Chambon.
- Van Fraassen, B.C. 1984, “Belief and the Will”, *Journal of Philosophy*.
- Widlöcher, D. 1994 “Remarques sur la croyance délirante”, *Revue Internationale de Psychopathologie*, 14
- Williams, B. 1973 “Deciding to Believe”, in *Problems of the self*, Cambridge, Cambridge University Press.

