

La satisfaction d'être dupe

Renée Bilodeau

Volume 28, numéro 2, automne 2001

URI : <https://id.erudit.org/iderudit/005663ar>

DOI : <https://doi.org/10.7202/005663ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Société de philosophie du Québec

ISSN

0316-2923 (imprimé)

1492-1391 (numérique)

[Découvrir la revue](#)

Citer cet article

Bilodeau, R. (2001). La satisfaction d'être dupe. *Philosophiques*, 28(2), 381–393.
<https://doi.org/10.7202/005663ar>

Résumé de l'article

Je me propose d'examiner la solution davidsonienne au problème de la duperie de soi afin de clarifier en quel sens il s'agit d'un acte intentionnel. Après une étude de quelques difficultés liées au concept même de duperie de soi, mon analyse met en lumière que la notion de partition de l'esprit que Davidson emprunte à son traitement de la faiblesse de la volonté ne peut être appliquée de manière satisfaisante à ce nouveau problème. J'indique ensuite que non seulement Davidson mais la majorité des philosophes qui étudient le phénomène ont négligé de donner un traitement adéquat du rôle qui est dévolu aux pro-attitudes dans la formation des croyances. Je reviens finalement à la question fondamentale, celle de savoir s'il est approprié de considérer que la duperie de soi est un acte intentionnel, afin de mettre en évidence le caractère paradoxal de tout raisonnement pratique qui prétendrait rationaliser un tel état ou processus.

La satisfaction d'être dupe

RENÉE BILODEAU

Faculté de philosophie

Université Laval

Renee.Bilodeau@fp.ulaval.ca

RÉSUMÉ. — Je me propose d'examiner la solution davidsonienne au problème de la duperie de soi afin de clarifier en quel sens il s'agit d'un acte intentionnel. Après une étude de quelques difficultés liées au concept même de duperie de soi, mon analyse met en lumière que la notion de partition de l'esprit que Davidson emprunte à son traitement de la faiblesse de la volonté ne peut être appliquée de manière satisfaisante à ce nouveau problème. J'indique ensuite que non seulement Davidson mais la majorité des philosophes qui étudient le phénomène ont négligé de donner un traitement adéquat du rôle qui est dévolu aux pro-attitudes dans la formation des croyances. Je reviens finalement à la question fondamentale, celle de savoir s'il est approprié de considérer que la duperie de soi est un acte intentionnel, afin de mettre en évidence le caractère paradoxal de tout raisonnement pratique qui prétendrait rationaliser un tel état ou processus.

ABSTRACT. — This paper discusses Davidson's solution to the problem of self-deception in order to clarify whether self-deception can be considered as an intentional act. I first examine some difficulties the concept of self-deception gives rise to, and then show that the notion of a partition of the mind Davidson borrows from his analysis of akrasia cannot satisfactorily be applied to self-deception. Afterwards, I indicate that not only Davidson but most of the philosophers interested in self-deception have neglected to give an adequate account of the role of pro-attitudes in the formation of justified belief. Finally, I come back to my central question and argue that any practical reasoning purported to rationalize such a state or process would be self-defeating.

Parmi les paradoxes de l'irrationalité, la duperie de soi est à ce point confondante que même les conditions nécessaires et suffisantes à une caractérisation appropriée du phénomène échappent encore au moindre consensus. De l'avis de certains, la duperie de soi doit être abordée par analogie avec la duperie d'autrui. Dans cette optique, la duperie de soi se présente comme une relation entre un trompeur et un dupe, le trompeur cherchant à amener sa victime à croire une proposition fausse ou non justifiée. Appliqué à la duperie interpersonnelle, ce canevas a peut-être quelque chose de grossier. Il est néanmoins, selon toute vraisemblance, fondamentalement adéquat. Appliquée à la duperie intrapersonnelle cependant, semblable façon de poser le problème engendre une série de difficultés ayant chacune leurs répercussions sur la notion de duperie de soi.

Avant d'entrer dans le vif du sujet, notons que j'ai introduit la notion de duperie de soi comme s'il s'agissait d'un processus. En fait, l'étiquette « duperie de soi » peut s'appliquer tant au processus par lequel un agent parvient à se tromper lui-même qu'à l'état qui résulte de ce processus. La chose

importante ici est que l'agent ne doit, à aucun moment, oublier qu'il croit que p , est enclin à croire que p , ou quelque chose d'analogue, aussi bien pendant qu'il s'emploie à acquérir la croyance que $\neg p$ que lorsqu'il en vient à croire que $\neg p$ ¹. Par exemple, si je range sur la plus haute tablette d'un placard où je ne vais jamais une potiche que j'abhorre, mais dont je n'ose me débarrasser parce qu'elle m'a été donnée par ma belle-mère, me fiant à ma mauvaise mémoire pour oublier l'endroit exact où je l'ai placée, je pourrai éventuellement réussir à acquérir une croyance fautive à propos de cette potiche. Il ne sera toutefois pas question de duperie de soi comme telle puisque la croyance dérangeante aura complètement disparu. Or même ceux qui n'entérinent pas le modèle interpersonnel réservent le nom de duperie de soi pour ces cas où l'individu doit s'employer avec persistance à maintenir à l'écart la croyance indésirable ou l'idée que les données empiriques supportent une proposition autre que celle qu'il croit. Quelle que soit la manière dont il les dissocie, l'agent doit donc avoir à l'esprit tant p que les propositions qui s'y opposent, qu'il s'agisse directement de $\neg p$, ou indirectement des propositions qui l'appuient.

Une interprétation stricte du modèle interpersonnel en a cependant conduit d'aucuns à aborder le phénomène de la duperie de soi avec scepticisme. On a ainsi fait valoir que, bien que cet état ou processus puisse paraître, de prime abord, empiriquement anodin, il s'avérerait après réflexion conceptuellement indéfendable. Afin de se duper elle-même, une personne devrait en effet tenir les deux rôles définis dans le scénario interpersonnel, celui du trompeur et celui du dupe. Elle devrait intentionnellement s'amener à croire la négation d'une proposition qu'elle entérine, tout en étant parfaitement consciente du jeu qu'elle joue avec elle-même. Mais comment rendre compte du fait qu'une personne S , qui croit que p , une proposition qui lui semble épistémiquement justifiée, puisse s'induire elle-même en erreur et en venir à répudier p au bénéfice de $\neg p$? Pire, lorsque le processus réussit, S tenant à la fois le rôle du trompeur et du trompé, comment imaginer qu'elle puisse croire que $\neg p$ tout en croyant que p ? Le modèle interpersonnel donne ainsi lieu à un paradoxe qui rendrait impossible la duperie de soi.

Diverses solutions ont été avancées, qui ont pour objet tantôt de concilier les éléments de ce paradoxe tantôt de montrer qu'il n'est qu'un artifice philosophique. Certains, tels David Pears ou Donald Davidson, ont mis de l'avant la thèse d'une partition de l'esprit, laquelle laisse place à la duperie de soi en alléguant que celui qui se dupe lui-même peut donner son assentiment à deux propositions contradictoires parce qu'il les relègue dans des sous-structures semi-autonomes de son esprit. D'autres, comme Alfred Mele ou Mark Johnston, défendent l'idée selon laquelle le modèle interpersonnel doit

1. Cette condition rend le processus et l'état de duperie de soi extrêmement voisins. Je ferai donc abstraction, dans les discussions qui suivent, des nuances existant entre les deux phénomènes.

être abandonné et insistent sur le fait que toute perspective intentionnelle sur la duperie de soi est une méprise.

Dans les pages qui suivent, je voudrais examiner la solution davidsonienne au problème de la duperie de soi afin de clarifier en quel sens il s'agit d'un phénomène intentionnel, si tant est qu'elle puisse être considérée tel. À mon avis, cette question est un défi lancé non seulement à ceux qu'intéresse le thème de la duperie de soi, mais à tous ceux qui travaillent dans le cadre du modèle causal d'explication de l'action. Après avoir brièvement exposé les parallèles existant entre la manière dont Davidson aborde différentes manifestations d'irrationalité, je présente et discute les principaux éléments autour desquels s'articule sa réponse à la question qui m'occupe. Je montre, dans un premier temps, que la notion de partition de l'esprit dont il s'était servi pour expliquer les actions akratiques ne semble guère qu'un procédé *ad hoc* lorsqu'elle est appliquée à la duperie de soi. J'indique ensuite que non seulement Davidson mais la majorité des philosophes qui étudient la duperie de soi ont négligé de donner un traitement adéquat du rôle qui est dévolu aux pro-attitudes dans la formation des croyances. Je conclus en revenant à la question fondamentale, celle de savoir s'il est approprié de considérer que la duperie de soi est un acte intentionnel, afin de mettre en évidence le caractère paradoxal de tout raisonnement pratique qui prétendrait rationaliser un tel état ou processus.

1. Arrière-fond théorique

Selon la théorie causale de l'action, un agent *S* accomplit intentionnellement une action si cette action est causée de la bonne façon par une raison, une raison étant minimalement composée d'une croyance et d'une pro-attitude. Ceci signifie deux choses : d'une part, qu'il existe un lien causal entre les actions intentionnelles et les événements mentaux dont elles sont l'effet ; et, d'autre part, que toute action intentionnelle peut être décrite comme si elle était le résultat d'un raisonnement pratique dont une croyance et une pro-attitude de l'agent seraient les prémisses. Cette deuxième remarque n'équivaut pas à l'affirmation que toute action intentionnelle est effectivement précédée par un raisonnement pratique explicite. Néanmoins, lorsque l'on classe un événement parmi les actions intentionnelles, il doit être possible de reconstruire un raisonnement plausible allant d'une croyance et d'une pro-attitude de l'agent à son intention d'agir. Ce cadre général, d'abord mis en place pour rendre compte des cas dans lesquels un agent accomplit une action, ou forme l'intention d'accomplir une action, en accord avec son meilleur jugement, peut être développé de manière à couvrir les cas dans lesquels un agent échoue soit à tirer la bonne conclusion de son raisonnement pratique, soit à s'y conformer. Ainsi, Davidson applique ce schéma aux actions akratiques et les présente comme des actions intentionnelles accom-

plies sur la base d'un raisonnement pratique dont la conclusion ne tient pas compte de toutes les données pertinentes. Une personne fait preuve de faiblesse de la volonté quand, dans une certaine situation, elle ne parvient pas à respecter ce que Davidson appelle le *principe de continence*. Selon ce principe, un agent devrait agir, ou former l'intention d'agir, en conformité avec ce qu'il juge être la meilleure option qui lui soit offerte, tout bien considéré. À titre d'exemple, considérons le cas de Jean, un fumeur invétéré. Depuis quelques semaines, Jean passe la plus grande partie de ses journées devant son ordinateur à écrire un livre qu'il espère terminer avant la fin de l'année académique. Jean traverse des moments particulièrement stressants et il a besoin de détente. Or, du moins le croit-il, le fait de fumer lui procure cette détente, en plus de l'aider à se concentrer sur ce qu'il écrit. Envisagée de la sorte, l'action de fumer apparaît à Jean comme justifiée par d'excellentes raisons. D'un autre côté, Jean a aussi de très bonnes raisons de cesser de fumer. Il sait que fumer est nocif pour la santé et accroît les risques de cancer. Qui plus est, Jean est déterminé à faire tout ce qui est en son pouvoir pour améliorer sa condition physique afin de contrer la fatigue provoquée par les longues heures de travail auxquelles il s'astreint. Ce second ensemble de considérations l'incline à penser qu'il devrait cesser de fumer. Après avoir pesé le pour et le contre, Jean en vient à conclure qu'il serait préférable de cesser de fumer. Et pourtant, il s'allume une cigarette...

En s'allumant une cigarette, Jean fait preuve de faiblesse de la volonté. Ceci n'empêche toutefois nullement que son action akratique soit une action intentionnelle typique. Elle est causée de façon non déviante par le désir qu'a Jean de se détendre et par sa croyance que fumer lui apportera la détente espérée, ce désir et cette croyance pouvant servir de prémisses dans le raisonnement pratique ayant pour conclusion l'intention ou l'action de fumer. Il s'agit néanmoins d'une action irrationnelle, car Jean ne respecte pas le principe de continence en regard duquel il devrait accomplir l'action qu'il a jugée la meilleure, en l'occurrence celle de cesser de fumer. Comment est-il possible qu'un agent opte délibérément pour une certaine action aux dépens d'une autre qu'il juge meilleure et qu'il pourrait tout aussi bien accomplir ? De l'avis de Davidson, une telle chose advient lorsqu'une partie de l'esprit se dissocie du tout et impose son propre désir à l'ensemble du système. En pareil cas, pour reprendre l'expression consacrée, « l'action est le résultat d'une cause mentale qui n'est pas une raison »². Ce n'est pas que l'agent n'ait aucune raison d'accomplir l'action qu'il accomplit, mais plutôt que la croyance et la pro-attitude qui causent alors l'action le fassent en dépit de la raison que l'agent juge la meilleure tout bien considéré.

2. Voir Davidson, 1970, p. 41. La distinction entre une raison et une cause mentale est reprise dans les différents textes de Davidson portant sur l'irrationalité. Il l'a récemment réitérée dans 1999, p. 404.

2. La duperie de soi

Dans quelle mesure ces remarques sont-elles susceptibles de nous aider à comprendre ce qu'il en est de la duperie de soi ? Afin de rendre plus clair le point en litige, considérons à nouveau un exemple. André, un étudiant de premier cycle en philosophie, a passé l'examen final de son cours de logique la semaine dernière. André sait qu'il a déjà échoué chacun des trois examens de logique précédents avec des résultats peu enviables. De plus, le calcul des prédicats du premier ordre est à ce point étranger à son esprit qu'il pourrait difficilement résumer la matière qui a fait l'objet des leçons des dernières semaines. André ne sait pas encore s'il a échoué ou non l'examen final, mais il en sait suffisamment sur ses piètres performances en logique pour ne pas se sentir très à l'aise lorsqu'il songe au résultat qu'il est susceptible d'obtenir. Néanmoins, parce qu'il a consacré beaucoup de temps à étudier en vue de son examen, André reste confiant et pense qu'il obtiendra finalement une note acceptable. En se rappelant toute l'énergie qu'il a mise dans la préparation de cet examen, André en arrive même à croire que sa note sera à tel point supérieure aux trois autres qu'elle compensera pour les échecs précédents et qu'il n'aura pas à reprendre le cours de logique.

La croyance d'André qu'il passera son cours de logique paraît être un cas typique de duperie de soi. Pourquoi en est-il ainsi ? Comme point de départ de la discussion, nous pouvons reprendre l'idée de Davidson selon laquelle la duperie de soi est un état ou processus qui met en jeu tout au moins les ingrédients suivants. *S* a une attitude cognitive envers une proposition *p*, cette attitude étant une croyance, une supposition, ou quelque autre attitude cognitive plus faible, pourvu que *S* croie *p* plus probable que sa négation. Pour une raison ou pour une autre, *S* trouve difficile à supporter l'idée que *p* soit vraie (ou soit probablement vraie). Cette idée le rend malheureux, anxieux, elle lui déplaît d'une quelconque manière³, et *S* préférerait que $\neg p$ advienne ou se soit produit. Afin d'éviter d'être confronté à l'idée que *p*, *S* acquiert la croyance que $\neg p$. *S* peut obtenir ce résultat par le biais de différentes stratégies : en réévaluant les données de manière tendancieuse (par un processus de rationalisation) ; en évitant de penser aux éléments perturbateurs (par l'évasion) ; en donnant une importance exagérée à des détails (par la surcompensation) ; en portant attention aux points qui lui conviennent aux dépens des autres ; etc.

Tous les auteurs qui admettent la possibilité de la duperie de soi s'entendent pour dire que de telles stratégies de manipulation des données sont à l'œuvre en pareil cas. Ils s'entendent également sur le fait que le rôle actif joué par l'agent qui trafique les informations est ce qui distingue la

3. Notez que celui qui se dupe lui-même ne se retrouve pas toujours dans un état dans lequel il se sent mieux qu'auparavant. Un mari jaloux peut se duper lui-même en acquérant la croyance que sa femme a un amant. Je ne chercherai pas à expliquer quels processus irrationnels amènent ce genre d'attitude.

duperie de soi d'autres formes d'irrationalité, notamment de la pensée magique (*wishful thinking*), avec laquelle elle partage plusieurs traits (en particulier, l'une et l'autre sont des croyances en une proposition mal fondée aux yeux de *S*, et l'une et l'autre sont des croyances provoquées par un état émotionnel). Néanmoins, les désaccords sont profonds en ce qui concerne la condition exacte dans laquelle se trouve la personne qui se dupe elle-même. Selon Davidson, ce pour quoi la duperie de soi est tellement intrigante tient au fait que celui qui se dupe lui-même croit que *p*, croit que $\neg p$, mais sans croire que *p* et $\neg p$, c'est-à-dire sans additionner 2 et 2. Pire, celui qui se dupe lui-même en vient à croire que $\neg p$ précisément parce qu'il croit que *p*. Si de nouvelles données empiriques recueillies de façon non biaisée par *S* avaient pour conséquence que la croyance que *p* cessait de lui apparaître épistémiquement justifiée, il pourrait abandonner cette croyance et entériner la croyance que $\neg p$ sans se duper lui-même. Mais dans l'esprit de celui qui se dupe, les deux croyances coexistent, et la croyance que *p* soutient la croyance que $\neg p$ ⁴. Pareille conception de la duperie de soi lui donne une allure nettement paradoxale qui en fait un défi particulièrement intéressant pour l'analyse philosophique. De l'opinion de Davidson, cette description de la duperie de soi est même la seule manière de résoudre les difficultés philosophiques qu'elle soulève sans les dissoudre. De l'avis des opposants à Davidson, toutefois, il s'agit d'une mise en scène qui intellectualise la question à outrance et qui est empiriquement (psychologiquement) inadéquate. Pour ces raisons, loin de contribuer à faire progresser les réflexions sur le problème, elle constitue un obstacle majeur à sa solution.

3. La position de Davidson

Qu'en est-il exactement de la façon dont Davidson parvient à expliquer la possibilité d'un état psychologique aussi contradictoire que la duperie de soi telle qu'il l'a définie ? La duperie de soi étant un phénomène de formation (ou de maintien⁵) d'une croyance biaisée, il faut d'abord examiner les situations où la formation de croyance se fait selon les règles pour être en mesure de saisir quels mécanismes font défaut quand une personne se dupe. Dans les cas les plus simples de formation de croyance, *S* se trouve dans un contexte dans lequel elle dispose d'un ensemble de données dont certaines appuient *p* tandis que d'autres appuient $\neg p$ ⁶. À la lumière de toutes les informations pertinentes, *S* évalue si *p* est plus probable que sa négation. Après avoir

4. Voir Davidson, 1985, p. 145, pour un résumé de sa façon de concevoir le problème.

5. Pour ne pas alourdir l'exposé, je m'exprimerai dorénavant uniquement en termes de formation de croyance.

6. Le cas limite où toutes les données vont dans le même sens ne peut servir à clarifier la difficulté que nous affrontons ici.

arrêté son jugement, *S* applique ce que Davidson appelle, à l'instar de Rudolf Carnap et Carl G. Hempel, le réquisit des données globales pour le raisonnement inductif (*requirement of total evidence for inductive reasoning*, par la suite RDG), un principe normatif qui enjoint d'accorder crédibilité à la proposition qui est le mieux supportée par toutes les données pertinentes dont dispose un individu⁷. Quand *S* échoue dans l'application du RDG, il est coupable de ce que Davidson appelle la « faiblesse de la justification » (*weakness of the warrant*). La faiblesse de la justification est une forme d'irrationalité dont la structure est analogue à celle de la faiblesse de la volonté, l'agent péchant ici contre le RDG, alors qu'il est en faute parce qu'il n'applique pas le principe de continence quand il fait preuve de faiblesse de la volonté. Être coupable de faiblesse de la justification n'est pas équivalent à se duper soi-même, mais cela en constitue une condition nécessaire. *S* ne peut se duper lui-même à moins qu'il ne respecte *habituellement* le RDG et échoue, dans un cas particulier, à l'appliquer. Ainsi, pour revenir à notre exemple, nous pouvons dire qu'André se dupe lui-même parce que, entre autres choses, il juge que l'ensemble des données dont il dispose appuie la proposition selon laquelle il va échouer son cours de logique, mais qu'il accorde crédibilité à la proposition, épistémiquement non justifiée selon lui, qu'il passera son cours⁸. Sa croyance ne cadre pas avec son évaluation des données parce qu'il ne respecte pas le RDG.

Que manque-t-il à la faiblesse de la justification pour tomber sous le coup de la duperie de soi ? Comme je l'ai mentionné précédemment, les éléments essentiels à la duperie de soi sont que celui qui se dupe lui-même croit que *p*, croit que $\neg p$, ne croit pas que *p* et $\neg p$, et croit que $\neg p$ parce qu'il croit que *p*. Le non-respect du RDG permet l'acquisition de la croyance que $\neg p$, mais ceci ne constitue que l'aspect cognitif du problème. En affirmant que *S* croit que $\neg p$ parce qu'il croit que *p*, nous introduisons dans le tableau une dimension motivationnelle qui témoigne d'un double aspect par lequel duperie de soi et faiblesse de la justification se distinguent. Il n'est question de duperie de soi, d'une part, que si *S* tient un rôle actif dans le processus qui a pour effet qu'il s'induit à la faiblesse de la justification. *S* ne se rend pas coupable de faiblesse de la justification par distraction ou à la suite d'une erreur de calcul. Il manigance pour se tromper lui-même, bien qu'il ne se représente

Par ailleurs, je laisse pour l'instant de côté la rationalité prudentielle, sur laquelle je reviendrai brièvement à la fin de ce texte. Une raison prudentielle de croire que *p* est une raison d'acquiescer ou d'entretenir une croyance en la proposition *p* pour les bienfaits pratiques que procure cette croyance. Les raisons prudentielles doivent être distinguées des raisons épistémiques de croire que *p*, c'est-à-dire des données empiriques à la lumière desquelles on peut accepter la vérité de *p*. Le portrait que je donne de la formation de croyance s'en tient à ces dernières. (Sur cette distinction, voir Davidson, 1985, p. 143 ; Brian McLaughlin, 1988, p. 46 *sqq.*)

7. Voir Davidson, 1985, p. 140.

8. McLaughlin, 1988, p. 40 et 51 défend une position similaire.

évidemment pas la situation en ces termes. D'autre part, et en continuité avec le premier point, puisque *S* agit à dessein lorsqu'il cherche à acquérir une croyance en une proposition qu'il juge fausse ou non justifiée, il doit avoir une raison de se comporter de la sorte. La raison pour laquelle *S* croit que $\neg p$ n'est pas d'ordre épistémique, elle ne relève pas des données empiriques ainsi qu'il en est, typiquement, des croyances bien formées. Sa raison est plutôt du même ordre que celles qui donnent lieu aux actions intentionnelles. Elle comprend, minimalement, une croyance et une pro-attitude, lesquelles expliquent l'acquisition de la croyance que $\neg p$. En disant que *S* croit que $\neg p$ parce qu'il croit que *p*, nous signifions que la croyance indésirable que *p*, ou les propositions qui l'appuient, fournissent à *S* le motif pour lequel il s'induit à la faiblesse de la justification. Le fait que « croit que *p* » soit imbriqué dans l'opérateur « parce que » ne doit pas donner à penser que la croyance que *p* rationalise l'action, comme le désir de se détendre, et la croyance que boire un verre de vin rouge est un bon moyen d'atteindre cette fin, rationalisent l'action de boire un verre de rouge. La croyance que *p* constitue un motif de croire que $\neg p$ dans l'unique mesure où l'agent souhaite que les choses soient différentes⁹. Il importe d'ailleurs de noter que ce que l'agent souhaite autre n'est pas le contenu de sa croyance, mais bien l'état du monde. La croyance que *p* conduit *S* à former la croyance que $\neg p$ parce que *S* désire que $\neg p$. La pro-attitude qui anime *S* n'est pas le désir de croire que $\neg p$. *S* ne souhaite pas acquérir une croyance dont le contenu serait $\neg p$, il souhaite qu'un certain état de choses se produise ou soit arrivé.

Ces précisions apportées, il est temps de revenir à notre question initiale, celle de savoir comment Davidson s'y prend afin de rendre compte de la possibilité de la duperie de soi. Pour une bonne part, la façon dont il procède s'inspire d'éléments introduits dans le but de traiter le problème de la faiblesse de la volonté. Dans un cas comme dans l'autre, l'existence au sein de l'esprit de structures semi-autonomes est une des clefs de voûte de son analyse. La faiblesse de la volonté est possible lorsqu'une partition de l'esprit advient, laquelle permet que les désirs d'une partie de celui-ci réussissent à outrepasser les préférences du tout. La partition favorise également la duperie de soi, en autorisant cette fois deux choses : d'abord, que le RDG soit mis à l'écart de façon à devenir inopérant ; ensuite, que la croyance importune soit formée dans une certaine partie de l'esprit de l'agent et isolée des croyances du tout de manière à ne pas entrer en conflit avec la croyance favorite qui occupe l'avant-plan. La cause de la partition est bien sûr le désir que la proposition que le RDG commande de croire soit fausse. Et quoique ce désir soit un événement mental, il s'agit d'une cause (mentale) qui n'est pas une raison. Aucun désir, aucune croyance ne sauraient justifier qu'un individu n'accorde pas crédibilité à la proposition qui est le mieux supportée par l'ensemble des données pertinentes dont il dispose. Qui plus est, entre la croyance de

9. Exception faite de certaines bizarreries déjà mentionnées (note 3).

prédilection de l'agent et la croyance indésirable, nul lien autre que causal ne saurait exister. Les deux croyances ne peuvent figurer dans un même raisonnement puisqu'elles appartiennent à des régions différentes de l'esprit. Le caractère irrationnel de la duperie de soi dépend directement de ce fractionnement de l'esprit auquel ont conduit certaines manœuvres douteuses de l'agent.

4. Critique du modèle davidsonien

Cette manière de résoudre la difficulté laisse perplexe à plus d'un égard. L'hypothèse voulant qu'une partition de l'esprit permette à des croyances contradictoires de coexister a peut-être de quoi séduire, mais ainsi introduite sans autre argument que celui de la nécessité, elle a des allures *ad hoc* qui la rendent suspecte. Davidson, il est vrai, indique sommairement quels mécanismes conduisent à cette partition et quel état les déclenche. Il contourne ainsi, jusqu'à un certain point, les accusations d'inadéquation psychologique souvent portées contre sa théorie. Néanmoins, faire valoir que la partition est opérée grâce à diverses stratégies de manipulation des données causées par le désir qu'un état de choses pire qu'incertain advienne ne s'avère une explication satisfaisante ni de l'existence de la partition, ni du bien-fondé de la caractérisation davidsonienne de la duperie de soi. Manipuler les données pourrait aussi bien avoir comme effet, ainsi que le soutient Mele¹⁰, que *S* ne possède jamais la croyance dérangeante que *p*, qu'elle s'empresse plutôt d'acquérir aussi rapidement que possible la croyance que $\neg p$. Une telle manœuvre serait réalisable pourvu que *S* s'emploie à tronquer les données dès qu'elle aperçoit qu'elles vont éventuellement appuyer *p* plutôt que $\neg p$, c'est-à-dire avant que *p* n'apparaisse épistémiquement plus probable que $\neg p$, et que *S* y croie.

Quiconque donne préséance à l'analyse de Mele sur celle de Davidson conviendra vraisemblablement avec celui-là que l'hypothèse d'une partition de l'esprit est inutile¹¹. En effet, même en supposant que le processus décrit par Mele n'est réalisable que si *S* néglige le fait que certaines informations appuient fortement *p*, il n'y a pas lieu d'introduire une hypothèse aussi forte que celle de la partition de l'esprit pour rendre compte de cette attitude. Qui plus est, tel n'est nullement le but dans lequel Davidson a mis de l'avant la notion de partition de l'esprit : la manipulation des données conditionne la partition de l'esprit, mais celle-ci a pour fonction de maintenir à l'écart le RDG d'une part, et la croyance importune d'autre part. En l'absence de la croyance que *p* et en regard de nombreux éléments appuyant $\neg p$, pourquoi serait-il besoin de compartimenter l'esprit de *S* ?

10. Voir Mele, 1987, chap. 9 ; 1997, p. 95.

11. Voir Mele, 1987, p. 75 *sqq.*

Outre le fait que la thèse de la partition de l'esprit ne puisse s'appuyer de façon concluante sur l'idée communément admise selon laquelle duperie de soi et manipulation des données vont de pair, l'affirmation même que pareilles stratégies sont à la base de la duperie de soi, pour intuitive qu'elle paraisse, entraîne plus de difficultés qu'elle n'en résout. En effet, la question que soulève immédiatement l'opinion d'après laquelle le sujet manipule les données est : comment un tel processus peut-il être initié et soutenu ? Tant que l'agent n'est pas dupe et qu'il croit, à tort ou à raison, que p est épistémiquement justifiée, comment peut-il engager une série de manœuvres pour en venir à croire que $\neg p$? Cette question peut être interprétée de deux façons. Elle peut simplement concerner les motifs qui incitent un agent à se duper lui-même, auquel cas elle sera résolue par des considérations psychologiques, en évoquant les désirs, les craintes, l'anxiété, etc., qui poussent l'agent à se débarrasser de la croyance dérangeante. Mais elle peut aussi être interprétée à un second niveau, à savoir : comment des désirs, craintes ou autres peuvent-ils avoir un impact prépondérant sur le contenu d'une croyance ou sur le processus menant à sa formation, c'est-à-dire sur des éléments d'ordre cognitif qui normalement devraient se fonder sur des données empiriques ou de nouvelles informations, non sur des souhaits ou des appréhensions ?

Les pro-attitudes d'un agent sont susceptibles de mettre en œuvre et de soutenir un processus cognitif de plusieurs façons, qui ne tombent pas toutes sous le coup de l'irrationalité. Une manière qu'ont les désirs d'engendrer une croyance peut être exemplifiée par le pari de Pascal. Pascal en vient à croire en Dieu parce qu'il le souhaite et s'y applique. Il sait qu'il a acquis cette croyance non sur la base de raisons épistémiques mais pour des raisons prudentielles, c'est-à-dire par intérêt pour les conséquences pratiques liées à l'acquisition de la croyance en Dieu. Il n'est cependant pas clair que le pari de Pascal soit un cas de duperie de soi, ne fut-ce que parce que Pascal cesse de croire que Dieu n'existe pas et de lutter contre sa croyance initiale. Et même s'il en était un, il est clair que tous les cas de duperie de soi ne reposent pas sur des raisons prudentielles.

Le pari de Pascal est un cas bien particulier, mais il en est d'autres dans lesquels un agent se trouvant en présence de données qui militent largement en faveur de p peut se mettre en quête de données appuyant $\neg p$ parce que l'idée que p est vraie lui déplaît sans pour autant se duper lui-même. Le type de pro-attitudes qui motive un agent ne saurait à lui seul servir de critère pour juger adéquatement des cas de duperie de soi. De plus, il ne peut être question de duperie de soi dès qu'un agent refuse de croire une proposition sur la base des premières données qui lui sont présentées. Finalement, et plus globalement, le problème qui se pose non seulement à Davidson mais aussi à Mele et à quelque autre philosophe favorable à l'idée, tout à fait défendable, que les pro-attitudes peuvent légitimement influencer sur la formation de croyance, est celui de préciser à quelles conditions ces dernières mènent à

l'acquisition d'une croyance épistémiquement justifiée plutôt que d'entraîner l'agent dans un processus de duperie de soi. Cette question voisine celle de savoir pourquoi certains désirs que $\neg p$ conduisent, en dépit des préférences de l'agent, à la croyance (justifiée) que p , alors que d'autres désirs de même contenu, dans des circonstances semblables, mènent à l'irrationalité¹². Cette manière de formuler la difficulté met en lumière le fait que tous les discours sur la manipulation des données, pour psychologiquement adéquats qu'ils sont, n'aident en rien à comprendre comment la duperie de soi est conceptuellement possible.

Ces remarques témoignent du fait, pourtant admis de tous, que la question philosophique fondamentale concerne la façon dont s'articulent les états mentaux de la personne coupable de duperie de soi. Seuls les liens entre croyances et pro-attitudes permettent d'évaluer l'intentionnalité et la rationalité d'une action. Ainsi que je l'ai rappelé en débutant ce texte, une action est intentionnelle lorsqu'elle est causée de la bonne façon par une raison, les raisons étant constituées minimalement d'une croyance et d'une pro-attitude. Pour en revenir à l'exemple d'André, il croit qu'il va échouer son cours de logique, cette pensée lui est pénible et il veut en être débarrassé. Ces états mentaux (entre autres) causent de manière non déviante l'acquisition de la croyance qu'il n'échouera pas. La formation de cette croyance sera considérée comme une action intentionnelle si les causes mentales dont elle découle sont aussi des raisons. Tel est le cas si nous pouvons rendre compte de l'acquisition de la croyance grâce à un raisonnement pratique que se tiendrait André¹³ :

- 1) Éviter la douleur est préférable (*ceteris paribus*).
 - 2) Penser que je vais échouer est douloureux.
- ∴ Il est préférable d'éviter de croire que je vais échouer, *ceteris paribus*.

Le raisonnement est valide, mais il n'a pas pour conséquence qu'André croit qu'il réussira son cours de logique. André doit d'abord former l'intention d'acquérir la croyance qu'il n'échouera pas, et ensuite se livrer à ces fameuses manœuvres de manipulation des données grâce auxquelles il parviendra éventuellement à se convaincre qu'il passera son examen comme il le souhaite. Si le stratagème donne les résultats escomptés, tant en raison de ses causes que dans la mesure où elle pourra être expliquée grâce à la construction d'un raisonnement pratique *bona fide*, l'acquisition de la croyance qu'il n'échouera pas ressemblera trait pour trait à l'accomplissement de n'importe quelle action intentionnelle. De la même façon que nous avons remarqué que la partition de l'esprit permettait d'affirmer que les actions akratiques sont des actions intentionnelles quoique irrationnelles, nous devons sembla-

12. Jeffrey Foss, 1997, p. 112 adresse une critique analogue à Mele.

13. Je reprends un raisonnement formulé par Davidson, 1985, p. 146 à l'appui de la thèse selon laquelle la duperie de soi est intentionnelle.

blement conclure que l'acquisition de la croyance que $\neg p$, bien qu'irrationnelle, est intentionnelle.

Le raisonnement prêté à André et sur la base duquel nous venons de juger que l'acquisition de la croyance que $\neg p$ est intentionnelle est-il cependant admissible ? Rien n'est moins sûr. En effet, ce raisonnement pose deux problèmes patents. D'une part, la présence d'une intention de former la croyance que $\neg p$ n'aide pas mais, à l'inverse, nuit à la formation de la croyance visée¹⁴. Tant que l'agent demeure conscient qu'il agit en vue de former une croyance non justifiée, il ne parviendra pas à acquérir cette croyance. Objecter que l'intention de l'agent n'est pas consciente ne fait que reculer le problème puisqu'il faudra alors expliquer comment cette intention a pu devenir si opportunément inconsciente. D'autre part, tel qu'il a été reconstruit, ce raisonnement laisse en veilleuse sa raison même d'exister, à savoir que, compte tenu de l'ensemble des données dont il dispose, André est persuadé qu'il va échouer. Or, souhaiter qu'une proposition soit fausse est une chose, parvenir à croire qu'elle l'est en se fondant sur un raisonnement dont le déroulement présuppose de manière non équivoque que les données prises en considération appuient au contraire la vérité de cette proposition en est une tout autre. À coup sûr, le raisonnement d'André porte en lui-même les causes de sa propre défaite.

Bibliographie

- Carnap, Rudolf, *Logical Foundations of Probability*, Chicago, University of Chicago Press, 1950, 2^e éd. revue et corrigée, 1962.
- Davidson, Donald, « How is Weakness of the Will Possible ? », dans *Moral Concepts*, Feinberg, Joel, dir., Oxford : Oxford University Press, 1970 ; repris dans *Essays on Actions and Events*, Oxford : Clarendon Press, 1980, 21-42.
- Davidson, Donald, « Paradoxes of Irrationality », dans Wollheim, Richard et Hopkins, James, dirs., *Philosophical Essays on Freud*, Cambridge : Cambridge University Press, 1982, 289-305.
- Davidson, Donald, « Deception and Division », dans LePore, Ernest et McLaughlin, Brian, dirs., *Actions and Events, Perspectives on the Philosophy of Donald Davidson*, Oxford : Basil Blackwell, 1985, 138-148.
- Davidson, Donald, « Reply to Ariela Lazar », dans Hahn, Lewis Edwin, dir., *The Philosophy of Donald Davidson*, The Library of Living Philosophers, Vol. XXVII, Chicago : Open Court, 1999, 402-405.

14. Ariela Lazar, 1999, p. 277 formule une objection similaire contre les théories intentionnelles de la duperie de soi.

- Foss, Jeffrey E., « How Many Beliefs Can Dance in the Head of the Self-Deceived ? » *Behavioral and Brain Sciences*, 20, 1997, 111-112.
- Hempel, Carl G., « Inductive Inconsistencies », dans *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York, The Free Press, 1965, 53-79.
- Johnston, Mark, « Self-Deception and the Nature of Mind », dans B. McLaughlin, A. Oksenberg Rorty, dirs., 1988, 63-91.
- Lazar, Ariela, « Deceiving Oneself Or Self-Deceived ? On the Formation of Beliefs « Under the Influence » », *Mind*, 108, 430, 1999, 265-290.
- McLaughlin, Brian, « Exploring the Possibility of Self-Deception in Belief », dans B. McLaughlin, A. Oksenberg Rorty, dirs., 1988, 29-62.
- McLaughlin, Brian, et Amélie Oksenberg Rorty, dirs., *Perspectives on Self-Deception*, Berkeley : University of California Press, 1988.
- Mele, Alfred, *Irrationality, An Essay on Akrasia, Self-Deception, and Self-Control*, New York : Oxford University Press, 1987.
- Mele, Alfred, « Real Self-Deception », *Behavioral and Brain Sciences*, 20, 1997, 91-102.
- Pears, David, « How Easy is Akrasia ? », *Philosophia*, 11, 1982, 33-50.
- Pears, David, *Motivated Irrationality*, Oxford : Clarendon Press, 1984.

