Document generated on 06/01/2025 4:16 a.m.

Philosophiques

Le dilemme de Bratman : problèmes de la rationalité dynamique

Paisley Livingston

Volume 20, Number 1, Spring 1993

URI: https://id.erudit.org/iderudit/027204ar
DOI: https://doi.org/10.7202/027204ar

See table of contents

Publisher(s)

Société de philosophie du Québec

ISSN

0316-2923 (print)
1492-1391 (digital)

Explore this journal

Cite this article

Livingston, P. (1993). Le dilemme de Bratman : problèmes de la rationalité dynamique. *Philosophiques*, *20*(1), 47–67. https://doi.org/10.7202/027204ar

Article abstract

This paper reconstructs the theory of dynamic rationality sketched by Michael Bratman in Intention, Plans, and Practical Reason. Evaluating agent's rationality is not just a matter of evaluating the agent's reasons for acting at the moment of a decision. We must ask not only whether it was rational of the agent to form a future-directed intention, but also whether it is rational of the agent to revise or retain that intention when the moment of action arrives. Such judgements are internal to the framework of the agent's ongoing plans, intentions, and the deliberative processes upstream of every decision. Yet they are also external because they refer to unintentional behavior and conditions beyond the agent's purposive control, such as habits and dispositions. I describe the dilemma Bratmans theory attempts to resolve, present Bratmans normative principles, and identify some outstanding problems. I propose a revision of one of Bratmans principles, arguing for a counterfactual evaluation of non-intentional factors as opposed to his consequentialist approach based on the long-term benefits of the agent's habits.

Tous droits réservés © Société de philosophie du Québec, 1993

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

PHILOSOPHIQUES, VOL. XX, NUMÉRO 1, PRINTEMPS 1993, P. 47-67

LE DILEMME DE BRATMAN : PROBLÈMES DE LA RATIONALITÉ DYNAMIQUE

par

Paisley Livingston

RÉSUMÉ : *Cet article propose une reconstruction de la théorie de la rationalité dynamique esquissée par Michael Bratman dans* Intention, Plans, and Practical Reason. *Évaluer* la rationalité de l'agent, *dit Bratman, ce n'est pas simplement évaluer les raisons d'agir qu'avait l'agent au moment de sa décision. Il faut se demander non seulement si l'agent était rationnel lorsqu'il a formé son intention d'agir, mais aussi s'il l'était encore en gardant ou en abandonnant cette même intention. Il s'agit d'une perspective* interne *dans la mesure où elle prend en compte les rapports entre l'action, l'intention qui la guide et le processus de délibération qui est en amont de celle-ci. Mais la perspective que nous adoptons lorsque nous évaluons la rationalité de l'agent est aussi* externe, *car elle fait référence à des conditions dont l'agent n'a pas la maîtrise, comme les dispositions et les habitudes qui lui sont propres et qui se manifestent dans ses comportements non intentionnels. Dans un premier temps, je décris le dilemme que la théorie de Bratman cherche à résoudre ; dans un second temps, je présente les principes que Bratman développe; finalement, j'en évalue les résultats en identifiant certains des problèmes qui restent à résoudre; je propose de réviser l'un des principes de Bratman en substituant une approche contrefactuelle à l'approche conséquentialiste qu'il propose.*

ABSTRACT : *This paper reconstructs the theory of dynamic rationality sketched by Michael Bratman in* Intention, Plans, and Practical Reason. *Evaluating* agent's rationality *is not just a matter of evaluating the agent's reasons for acting at the moment of a decision. We must ask not only whether it was rational of the agent to form a future-directed intention, but also whether it is rational of the agent to revise or retain that intention when the moment of action arrives. Such judgements are* internal *to the framework of the agent's ongoing plans, intentions, and the deliberative processes upstream of*

every decision. Yet they are also external *because they refer to unintentional behavior and conditions beyond the agent's purposive control, such as habits and dispositions. I describe the dilemma Bratman's theory attempts to resolve, present Bratman's normative principles, and identify some outstanding problems. I propose a revision of one of Bratman's principles, arguing for a counterfactual evaluation of non-intentional factors as opposed to his consequentialist approach based on the long-term benefits of the agent's habits.*

L'une des thèses principales du livre de Bratman, *Intention, Plans, and Practical Reason* est que notre rationalité est une rationalité *située*[1]. Nous sommes situés, avant tout, dans le temps. Nous agissons et nous réfléchissons dans le présent, mais dans ce même temps, nous nous souvenons du passé et nous faisons des projets pour l'avenir. Ainsi, Bratman dit que nous ne sommes pas des « *time-slice agents* », c'est-à-dire de simples coupes synchroniques.

L'un des buts de Bratman est d'esquisser une théorie normative de la rationalité dynamique. Selon lui, deux perspectives s'ouvrent devant nous lorsque nous procédons à l'évaluation de la rationalité *d'une action*. Une première, dite *interne*, se situe à l'interieur du cadre constitué par les projets antérieurs qu'avait l'agent ; une deuxième, dite *externe*, met ces projets hors jeu et ne prend en compte que les attitudes de l'agent au moment même de son action. Parfois ces deux perspectives ne convergent pas sur une évaluation unique de l'action.

---

1. Michael E. Bratman, *Intention, Plans, and Practical Reason*, Cambridge, Mass., Harvard University Press, 1987. Voir aussi les articles de Bratman, « What is Intention ? », Stanford, *Center for the Study of Language and Information*, 1987 ; « Intention and Personal Policies », Stanford, *CSLI*, 1988 ; « Planning and the Stability of Intention », *Minds and Machines*, 2, 1992, p. 1-16 ; et « Practical Reasoning and Acceptance in a Context », *Mind*, 101, 1992, p. 1-15 ; aussi bien que Martha E. Pollack, David J. Israel, et Michael E. Bratman, « Toward an Architecture for Resource-Bounded Agents », Stanford, *CSLI*, 1987 ; Philip R. Cohen et Hector J. Levesque, « Persistence, Intention, and Commitment », Stanford, *CSLI*, 1987 ; et Edward F. McClennen, *Rationality and Dynamic Choice : Foundational Explorations*, Cambridge, Cambridge University Press, 1990.

Selon Bratman, ces deux perspectives échouent à saisir certains aspects de la rationalité dynamique. Il croit qu'il est nécessaire de concevoir une autre perspective sur la rationalité, à savoir ce qu'il nomme *la rationalité de l'agent* (*agent's rationality*). Il s'agit d'une perspective « interne » dans la mesure où elle prend en compte les rapports entre l'action, l'intention qui la guide et le processus de délibération qui est en amont de celle-ci. Mais la perspective que nous adoptons lorsque nous évaluons la rationalité de l'agent est aussi *externe*, car elle fait référence à des conditions dont l'agent n'a pas la maîtrise, par exemple : les dispositions et les habitudes qui lui sont propres et qui se manifestent dans ses comportements non intentionnels.

Évaluer la rationalité de l'agent, dit Bratman, ce n'est pas simplement évaluer les raisons d'agir qu'avait l'agent au moment de sa décision. Il faut se poser la question qui consiste à savoir non seulement si l'agent était rationnel lorsqu'il a formé son intention d'agir, mais aussi s'il l'était encore en gardant ou en abandonnant cette même intention. La théorie de la rationalité de l'agent cherche à décrire les normes qui gouvernent ce genre de processus de délibération prolongée.

Dans ce qui suit, je proposerai une reconstruction de cette théorie. Dans un premier temps, je décrirai le dilemme que la théorie de Bratman cherche à résoudre ; dans un second temps, je présenterai les principes que Bratman développe ; finalement, je tâcherai d'en évaluer les résultats ; j'identifierai certains des problèmes qui restent à résoudre et je proposerai de réviser l'un des principes proposés par Bratman.

1. Le dilemme du projet

Dans la théorie de Bratman, les intentions sont les éléments constitutifs des projets. L'intention est, selon lui, « une attitude distinctive qu'il ne faut pas confondre avec les désirs et les croyances ni réduire à ceux-ci » (p. 10). Il y a deux sortes d'intentions : celles qui concernent le moment présent et celles qui visent un moment à venir. L'une des thèses de Bratman est que les modèles de la décision axés uniquement sur les préférences et les croyances n'arrivent pas à décrire le rôle que jouent les

intentions futures[2]. Celles-ci ont une stabilité particulière : une fois formées, elles ont tendance à persister ; qui plus est, elles constituent un cadre pour d'autres délibérations en identifiant les problèmes qu'il faut résoudre et en délimitant les solutions admissibles. Selon Bratman, les intentions futures ont aussi une dimension volitive, car si une intention persiste jusqu'au moment où l'action était prévue, l'agent aura tendance à l'entreprendre, toutes choses étant égales par ailleurs. L'intention manifeste donc un type d'engagement à faire l'action qui n'est pas celui du désir, et il est impossible de la réduire à une combinaison quelconque de désirs et de croyances[3].

Selon Bratman, l'une des clefs de la rationalité humaine est notre capacité à nous engager à faire quelque chose dans l'avenir. Mais, par l'adoption de cette thèse, il est confronté au dilemme du projet (Bratman décrit à vrai dire un trilemme, mais je mettrai de côté le spectre, à mon avis inintéressant, d'une objection métaphysique basée sur une « action à distance »). D'un côté du dilemme, se trouve une objection contre l'idée même que des êtres rationnels puissent s'engager réellement pour l'avenir. Après tout, étant donné la non-omniscience de l'agent, il lui est impossible de prévoir les conditions futures, et beaucoup de ses intentions et de ses projets ne lui conviendront plus quand le moment de les mettre en action sera arrivé. Sous de telles conditions, s'engager d'une manière résolue amène l'agent à poser bon nombre de gestes irrationnels. Mais de l'autre côté du dilemme, nous sommes confrontés à l'objection suivante : si les intentions face à l'avenir peuvent bel et bien être abandonnées, la rationalité se réduit au raisonnement qu'opère l'agent au moment même d'agir. De deux choses l'une : faire des projets et s'engager pour l'avenir sont des formes dangereuses d'irrationalité, ou une

---

2. Ici la terminologie de Bratman se distingue d'autres emplois dans la littérature. Alfred Mele, par exemple, se sert du « plan » afin de faire référence aux contenus sémantiques des intentions ; voir son *Springs of Action : Understanding Intentional Behavior*, New-York, Oxford, 1991, p. 144. Dans ce qui suit, je ne ferai pas de distinction entre « plan », « projet », « dessein » et « intention » (pour l'avenir).

3. Je n'aborderai pas ici le problème épineux qui consiste à se demander si l'on peut vraiment démontrer que les intentions sont irréductibles aux désirs et croyances. Sur ce point, voir Mele, *Springs of Action*.

simple perte de temps. Dans les deux cas, les intentions futures ne contribuent en rien à notre rationalité pratique.

Face à ce dilemme, Bratman concède que nos prévisions et nos projets sont bel et bien faillibles, mais il persiste à croire que nous en avons besoin. Des agents ayant des capacités cognitives très limitées ne peuvent pas se permettre d'attendre le dernier moment avant de se demander ce qu'il faut faire. La sagesse classique nous enseigne que les agents sans projet en font les frais. La rationalité est du côté de la fourmi, non de la cigale. Pour le dire d'une manière plus contemporaine, le Parisien qui attend jusqu'à la fin du mois de juillet pour se demander comment passer ses vacances d'été aura un vrai problème. Les intentions ne nous dictent pas nos gestes futurs, mais elles contribuent d'une manière indispensable à la solution de nos problèmes de coordination, car elles orientent nos raisonnements et font partie d'une disposition rationnelle.

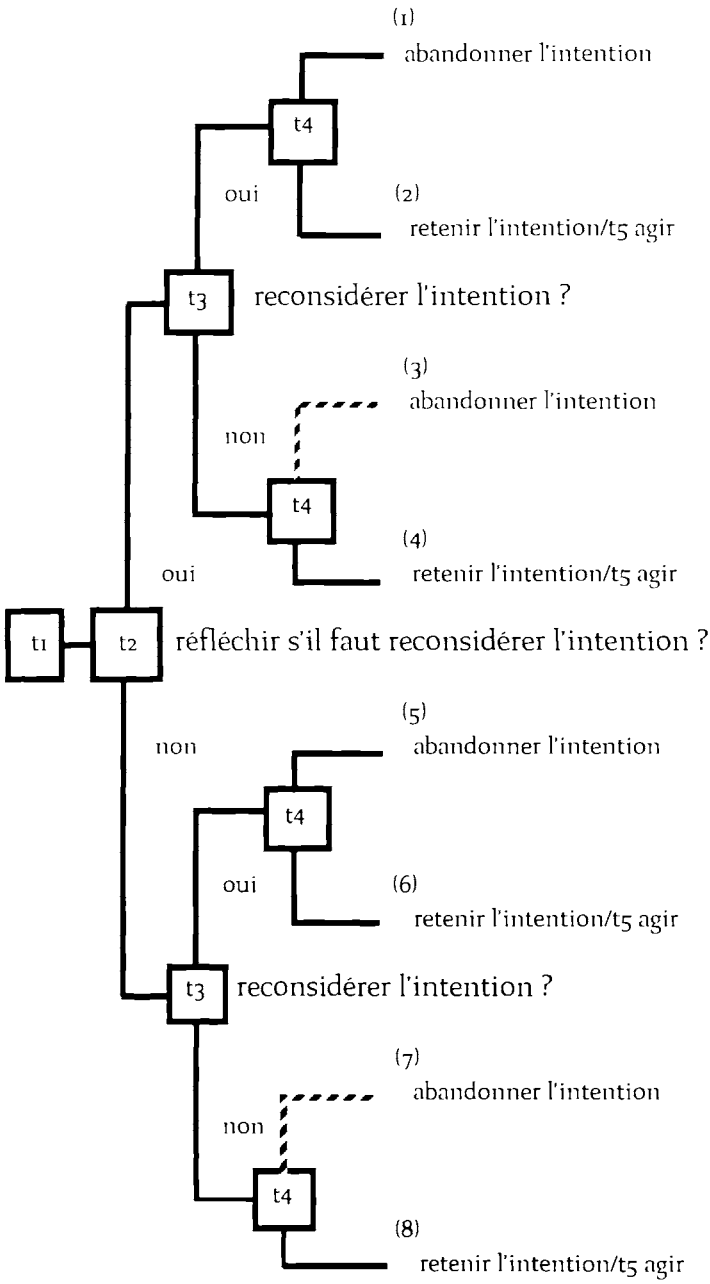
Telles sont, du moins, les intuitions de Bratman. Les principes de la rationalité des agents, qui en sont la codification théorique, se présentent comme les normes auxquelles doit satisfaire l'agent qui veut trouver une voie entre l'engagement néfaste et la pensée de dernière minute.

2. Les principes de la rationalité de l'agent

Afin d'identifier les éléments de la rationalité de l'agent, Bratman décrit un processus dynamique comportant quatre moments cruciaux, que nous pouvons nous représenter comme les nœuds d'un arbre de décision. Le processus en question comprend cinq moments :

1. Avant le moment t1, l'agent, S, n'a pas l'intention de faire une action, A, à un moment futur, t5. À t1, S se demande s'il ne devrait pas faire cette action, et le résultat de sa délibération conduit S à former l'intention de faire A à t5.
2. Pendant l'intervalle t2, soit S réfléchit sur la possibilité de reconsidérer l'intention formée à t1, soit S n'y réfléchit pas[4].

4. La distinction entre l'action de réfléchir sur la possibilité de reconsidérer ou non son intention et l'action même de la reconsidérer deviendra plus claire par la suite. Pour l'instant, imaginez la différence entre une pensée de second ordre du style « devrais-je rouvrir ce dossier ? » et le travail qui consiste à faire toute une



étude sur le bien-fondé d'un projet entier. La distinction étant faite, nous recontrons certes des cas mixtes où le fait de réfléchir comprend déjà un élément de reconsidération (voir Bratman, p. 97).

3. À t3, soit S reconsidère son intention, soit S ne la reconsidère pas.
4. À t4, soit S conserve son intention de faire A à t5, soit S l'abandonne.
5. Si à t4 l'intention a été conservée, S agit en fonction de cette intention, faisant A intentionnellement à t5.

(Consulter l'arbre de décision de la page ci-contre.)

Étant donné ce schéma, comment peut-on évaluer s'il est rationnel pour S de faire A ? En regardant la fin du processus, nous voulons savoir sous quelles conditions il est rationnel *pour un agent* de faire A à t5, et Bratman nous propose le principe suivant :

Principe de l'intention et de l'action

S'il est rationnel pour S d'avoir l'intention de faire A dans le présent, et si S réussit à accomplir cette intention, faisant A intentionnellement, alors il est rationnel pour S de faire A.

Mais si nous voulons appliquer ce principe, il nous faut savoir sous quelles conditions il est rationnel pour S d'avoir l'intention de faire A dans le présent, ce qui nous oblige à prendre en compte le processus qui a donné lieu à la formation de cette intention. Dans les modèles de la décision traditionnels, ce que la rationalité exige c'est que l'intention de faire A soit au moins aussi bien justifiée par les raisons de l'agent, c'est-à-dire par ses désirs et ses croyances, que tout autre choix qu'il envisage. Ce genre d'évaluation « atomique » du choix rationnel n'est pas le thème central des travaux de Bratman, et il ne cherche pas à prendre partie dans tous les débats qu'il peut y avoir au sujet de cette famille de modèles. Afin de se concentrer sur les problèmes spécifiques soulevés par les intentions et les projets, Bratman écarte ce qu'il appelle « les problèmes fondamentaux qui concernent l'évaluation de la rationalité d'un agent ayant tel ou tel désirs et croyances[5]. »

S'étant ainsi donné la possibilité d'évaluer la rationalité de la formation de l'intention à t1, nous sommes en mesure de regarder de plus près les moments t2 à t4, qui sont spécifiques à une théorie de la rationalité de l'agent. Si nous étudions notre arbre de décision, nous verrons qu'il y a huit chemins menant à deux

5. *Cf.* p. 58.

sortes de décision. S forme l'intention, il réfléchit ou non sur la possibilité de la reconsidérer, il la reconsidère ou pas, et enfin il garde l'intention ou l'abandonne. Selon une hypothèse provisoire et simplifiée, la rationalité de chaque nœud de l'arbre peut être jugée indépendamment. Deux valeurs sont possibles : soit rationnel, soit irrationnel ; ce qui nous donne 128 possibilités, dont certaines sont des cas de ce que Bratman appelle une « délibération étendue[6] ». Les chemins (3) et (7) impliquent l'abandon arbitraire d'un projet à la dernière minute, et ne seront pas considérés dans ce qui suit ; il nous reste donc 96 schèmes de délibération étendue. En voici un exemple :

DE1 S forme une intention rationnelle de faire A à t5 ; ensuite S réfléchit rationnellement sur l'opportunité de reconsidérer l'intention et S décide rationnellement de la reconsidérer. Puis sa décision, elle aussi rationnelle, est de conserver son intention initiale.

L'agent qui réalise ce scénario nous offre un exemple d'une rationalité d'agent bien réussie. Il faut opposer ce genre de cas à l'épisode suivant :

DE2 S forme une intention rationnelle de faire A à t5 ; ensuite ; d'une manière irrationnelle, S réfléchit sur l'opportunité de reconsidérer son intention et s'engage irrationnellement dans une reconsidération approfondie de l'intention. Le résultat en est que S décide finalement de retenir son intention initiale.

Ici l'agent finit par accomplir un acte rationnel, mais, sans le faire exprès, il gaspille son temps et ses ressources dans une reconsidération inutile. Réfléchissons aussi sur le genre de cas suivant :

DE3 S forme une intention irrationnelle de faire A à t5 ; ensuite S réfléchit rationnellement sur l'opportunité de reconsidérer l'intention et décide rationnellement de ne pas s'engager dans une reconsidération approfondie de l'intention. S retient son intention initiale.

---

6. *Cf.* p. 79. Je reviendrai sur la façon dont Bratman aborde le problème de l'évaluation de la rationalité de la (non-)reconsidération irréfléchie.

Bratman considère que de tels cas soulèvent un problème de « *bootstrapping* ». Voici pourquoi. Il est vrai, parfois, que nous aurions tort de nous arrêter en cours de route pour nous embourber dans une reconsidération détaillée de notre intention antérieure. Dans l'exemple que Bratman fournit, un homme politique est en train de débattre avec son adversaire à la télévision, et il voit que son rival vient de présenter un certain argument d'une manière particulièrement efficace. Ayant prévu d'attaquer son rival sur ce point même, l'homme politique se demande s'il est toujours opportun de suivre son dessein initial. Mais il comprend qu'il n'a pas le temps nécessaire à une reconsidération détaillée, et il se dit que s'il abandonnait son plan d'attaque initial, il risquerait une improvisation mal réussie. Après tout, il sait que son équipe a bien préparé le plan d'attaque et il connaît l'argument par cœur. Il poursuit donc le plan initial, tout en sachant que s'il avait eu le temps nécessaire, il l'aurait reconsidéré plus longuement. Sa non-reconsidération du projet résulte d'une réflexion rationnelle, mais celle-ci l'amène à conserver une intention irrationnelle. Il semblerait que l'irrationnel se soit converti en quelque chose de rationnel, ce qui est pour Bratman du « *bootstrapping* », qu'une bonne théorie de la rationalité de l'agent devrait exclure.

Bratman souligne donc les différences entre une évaluation interne de la rationalité d'une action et une évaluation de l'agent en train d'accomplir l'action. Dans le premier cas, nous prenons les intentions antérieures comme une donnée de base, ce qui veut dire qu'une intention qui était irrationnelle au moment de sa formation peut se convertir en une intention rationnelle sous certaines conditions — étant donné, certes, la perspective interne. Dans le deuxième cas, ce n'est pas seulement la rationalité de l'acte qui nous intéresse. Bratman distingue ce qu'il nomme une théorie « a-historique » d'une théorie « historique » de la rationalité de l'agent, tout en soulignant les avantages de cette dernière. Dans une perspective « a-historique », l'évaluation de la rationalité de l'agent qui a l'intention de faire quelque chose dans l'avenir ne dépend pas de l'histoire antérieure de cette intention. Le principe de base est que la rationalité de la non-reconsidération de l'intention implique la rationalité de l'agent. Une théorie dite historique de la rationalité de l'agent prend en compte, au contraire, deux sortes de faits, à savoir (1) la rationalité ou

l'irrationalité de la formation originale de l'intention ; et (2) la rationalité ou l'irrationalité de la non-reconsidération de l'intention. Les principes que nous propose Bratman visent ce deuxième problème.

Bratman avoue que proposer une théorie dans un tel domaine est une chose risquée : « Le problème est que si nous prenons au sérieux les considérations historiques nous risquons d'être submergés par leur complexité et leur hétérogénéité[7]. » Mais Bratman accepte ce risque et propose des principes de la rationalité historique de l'agent, dont voici le premier :

Principe de la rationalité délibérative

Si à t1 S forme l'intention de faire A à t5, et cela sur la base d'une délibération à t1, il est rationnel pour S à t1 d'avoir l'intention de faire A à t5 si et seulement si deux conditions sont remplies, à savoir :

(a) s'il y a des intentions qui jouent un rôle d'arrière-plan dans la délibération de S, il faut qu'il soit rationnel pour S de les avoir ; et

(b) à t4, S pense rationnellement que A est au moins aussi bien fondée que les autres options qui sont pour lui pertinentes et admissibles.

Comme je l'ai fait remarquer plus haut, Bratman ne cherche pas à expliquer les implications précises de la condition (b), car ce n'est pas un problème spécifique à une théorie de la rationalité dynamique. Notons aussi que l'application de la condition (a) amènera dans certains cas à une itération du même principe. Dans d'autres cas, nous arrivons à une ou à des intentions non-délibérative(s) (c'est-à-dire à des intentions qui ne furent pas reconsidérées par l'agent entre t1 et t5), et il nous faut une manière de les évaluer. À cette fin, Bratman propose un deuxième principe, que voici :

Principe de la rationalité non-délibérative

Il est rationnel pour S à t4 d'avoir l'intention de faire A à t5 si et seulement si deux conditions sont remplies, à savoir :

(a) il était rationnel pour S à t1 de former cette intention ; et

(b) il était rationnel pour S, de t1 à t4, de ne pas la reconsidérer.

---

7. *Cf.* p. 79.

La condition (a) est, encore une fois, une question concernant l'évaluation de la rationalité des désirs et des croyances de l'agent. Notre attention doit donc se porter sur la condition (b), qui soulève nombre de questions. Comment peut-on évaluer la rationalité de la non-reconsidération d'une intention ?

L'approche que propose Bratman est essentiellement « conséquentialiste », car il s'agit selon lui d'évaluer la propension de l'agent à reconsidérer ou non ses intentions premières, propension dont l'agent n'a peut-être pas la maîtrise. En effet, lorsqu'un agent reconsidère ou ne reconsidère pas une intention qu'il avait formée antérieurement, cela peut résulter soit d'un moment de réflexion, soit d'un geste irréfléchi. Dans les deux cas, c'est la manifestation d'une disposition propre à l'agent, qui peut dépendre, par exemple, de sa tendance à remarquer certains phénomènes particuliers, ou de sa propension à conserver un certain degré de confiance en ses propres projets et prévisions. Ici on peut penser au genre de certitude et d'incertitude que souligne André Orléan dans sa lecture de Keynes[8]. Pour Bratman, nous devons évaluer une telle disposition en analysant ses conséquences sur les intérêts à long terme de l'agent. Pour que la disposition soit jugée « raisonnable », il n'est pas nécessaire qu'elle conduise à une optimisation de l'utilité (intertemporelle) pour l'agent ; Bratman retient simplement un critère de « *satisficing* » valant pour le long terme. Ce seuil franchi, nous jugeons qu'il est « raisonnable » pour l'agent d'avoir la disposition en question. Présupposant que ce genre d'évaluation peut être réalisé (un point sur lequel je reviendrai plus tard), Bratman formule le principe suivant :

Principe de la rationalité de la (non-)reconsidération irréfléchie

La (non-)reconsidération irréfléchie d'une intention antérieure est rationnelle pour S si elle est la manifestation par S d'une disposition de reconsidération raisonnable.

8. Voir, par exemple, André Orléan, « Anticipations et conventions en situation d'incertitude », *Cahiers d'économie politique*, 13, 1987, p. 153-72 ; et « Pour une approche cognitive des conventions économiques », *Revue économique*, 40, 1989, p. 241-72.

Comment ce principe est censé s'appliquer deviendra plus clair dans un instant ; j'y apporterai des objections dans la dernière partie de mon texte.

Les principes de Bratman stipulent que les délibérations étendues d'un agent ne peuvent être rationnelles lorsqu'elles comprennent l'une des conditions suivantes : (a) une intention dont la formation fut irrationnelle ; (b) une reconsidération irrationnelle ; (c) une décision réfléchie irrationnelle de ne pas reconsidérer une intention antérieure ; ou (d) une disposition non raisonable de (non-)reconsidération. Si l'on se réfère maintenant à l'arbre de décision présenté plus haut, nous pouvons dire que les principes de Bratman servent à filtrer les 96 possibles scénarios de délibération, en excluant tous les cas où il y a un nœud irrationnel. Il peut même sembler que ce qu'il appelle la rationalité de l'agent n'est pas autre chose que le fait d'avoir la note « rationnel » sur une suite de quatre nœuds ; c'est de la rationalité atomique en brochette. Mais il faut aussi voir que les principes de Bratman admettent des cas où un aspect de la délibération étendue n'est pas parfaitement rationnel, car ce que Bratman appelle une disposition raisonnable peut donner lieu a des (non-) reconsidérations irréfléchies défectueuses. Prenons le cas suivant :

DE4 S forme une intention rationnelle de faire A à t5 ; à cause d'une disposition raisonnable, S ne réfléchit pas quant à l'opportunité de reconsidérer l'intention, et effectivement ne la reconsidère pas ; ensuite S retient (rationnellement) son intention initiale.

Il n'est pas du tout évident, dans de tels cas, qu'il soit toujours rationnel de retenir l'intention initiale, car il se peut que les circonstances aient changé radicalement entre t1 et t4. Le principe de Bratman n'exclut pas que, dans certains cas, la disposition raisonnable de l'agent le conduise à ne pas prendre note d'un tel changement de circonstances : si l'agent avait reconsidéré son intention, il l'aurait abandonnée. Dans ce cas, la rétention de l'intention par l'agent est irrationnelle vue d'une perspective externe, et il faut dire la même chose de l'action qui en résulte. Mais la disposition qui a donné lieu à ce comportement peut être globalement rationnelle, à condition simplement qu'elle serve assez bien l'utilité pour l'agent sur une longue période. Il faut donc conclure

qu'il y aura des cas (mais pas beaucoup) où une action pourra être jugée irrationnelle alors que l'agent qui l'accomplit pourra être jugé rationnel[9].

Considérons aussi le cas suivant :

DE5 S forme une intention rationnelle de faire A à t5 ; à cause d'une disposition raisonnable, S ne réfléchit pas sur l'opportunité de reconsidérer son intention. Mais la même disposition fait qu'il la reconsidère effectivement ; puis S décide rationnellement de retenir son intention initiale.

Notons ici que l'action est rationnelle, mais aussi que la reconsidération que fait S peut être une perte de temps. Cependant la disposition qui la cause est raisonnable, de telle sorte que nous voulons dire que l'agent est rationnel en accomplissant l'action, même s'il y a un élément d'irrationalité dans son comportement.

Les cas DE4 et DE5 nous montrent que les dispositions de reconsidération irréfléchie jugées globalement raisonnables n'excluent pas des éléments d'irrationalité, à savoir des reconsidérations inutiles ou des manques inappropriés de reconsidération. Ces deux genres de cas démontrent que les principes de la rationalité de l'agent que propose Bratman « pardonnent » à l'agent son incapacité de résoudre à chaque coup ce que nous avons nommé plus haut le dilemme du projet.

Considérons maintenant un autre cas :

DE6 S forme une intention rationnelle de faire A à t5 ; à cause d'une disposition raisonnable, S réfléchit ensuite sur l'opportunité de reconsidérer son intention et décide irrationnellement de ne pas s'engager dans une reconsidération approfondie de celle-ci, le résultat étant que S retient (rationnellement) son intention initiale.

Ici la rationalité de l'action doit se juger de la façon suivante : si l'agent avait reconsidéré rationnellement son intention et si, dans ce cas, il l'avait conservée, on aurait été conduit à dire que son intention et son action étaient rationnelles. Mais notons

9. Voir l'exemple de Bratman, p. 72-3, 86-7.

aussi que le processus de délibération comprend un moment de réflexion défectueuse qui, tout en étant le résultat d'une disposition raisonnable, demeure irrationnelle. Par exemple, en demandant s'il devait rouvrir le dossier, l'agent n'a pas tenu compte d'une bonne raison qu'il avait effectivement de le faire. Le principe historique de Bratman nous dit que cela ne peut pas être considéré comme un exemple de rationalité de l'agent, car il n'était pas rationnel pour celui-ci de ne pas reconsidérer l'intention antérieure.

Face à de tels cas, Bratman introduit une mise en garde. Nous ne pouvons pas toujours, dit-il, arriver à une évaluation binaire de rationalité ou d'irrationalité ; parfois il faut se contenter de dire que dans un certain sens l'agent n'est pas rationnel, mais qu'il n'est pas, non plus, irrationnel. Bratman considère que, dans de tels cas, il faut parler de la « non-rationalité » de l'agent. « C'est là le degré de précision que permet ce domaine » ajoute Bratman[10].

3. L'évaluation de la théorie

A-t-on résolu par le moyen de ces principes le dilemme du projet ? Cela dépend, certes, de ce que nous acceptons comme une solution véritable. Il faut admettre au moins que la théorie de Bratman identifie le genre de comportements que nous cherchons à cerner lorsque nous songeons à résoudre le dilemme : les agents doivent avoir la capacité de réfléchir sur et de reconsidérer non pas la totalité, mais certains de leurs projets et de leurs intentions à la lumière de l'évolution des circonstances. Les principes de Bratman nous aident à circonscrire cette capacité en proposant l'esquisse d'une analyse fonctionnelle.

Pourtant, l'on peut répliquer que les principes de Bratman ne nous disent pas *comment* un agent peut produire des délibérations étendues rationnelles et éviter celles qui sont irrationnelles. Ils ne nous disent pas non plus comment procéder à l'évaluation d'un épisode particulier de délibération. La théorie n'offre pas une description précise de ce qu'est cette disposition raisonnable de (non-)délibération. Comment dire si l'utilité (intertemporelle) est suffisamment améliorée par une certaine disposition pour que

10. *Cf.* p. 100.

celle-ci puisse être jugée raisonnable ? L'on pourrait ajouter qu'un faisceau de principes ne constitue pas, en tout cas, une véritable *théorie*, si par là on entend un algorithme capable d'engendrer les bons épisodes de délibération étendue.

Bratman pourrait répondre que la matière en question n'autorise pas un tel degré de précision, et que nos descriptions et nos évaluations de la rationalité de l'agent auront toujours une dimension idiographique irréductible. Le fardeau de la preuve incombe à ceux qui veulent réduire la *phronésis* à la *theôria*. Une fois que l'on a dit que la rationalité humaine est une rationalité située, ne faut-il pas admettre qu'il n'y a pas de véritable théorie de la rationalité dynamique ? Mais Bratman n'adopte pas cette position. Bien au contraire, car il considère que nous pouvons formuler des hypothèses générales au sujet des dispositions raisonnables, sans même avoir besoin de tenir compte de la nature des intérêts et des désirs qu'elles sont censées servir.

Quelles sont donc ces hypothèses sur la nature des dispositions raisonnables ? Bratman suggère qu'elles impliquent normalement une tendance à ne reconsidérer un de nos plans que quand nous sommes confrontés à un « problème pour ce plan », en distinguant « problème pour un plan » et « problème posé par un plan ». Tout plan pose des problèmes, car nos projets ne peuvent jamais spécifier en détail tout ce qu'il sera nécessaire de faire afin de les réaliser. Un conférencier a l'intention de présenter un texte la semaine prochaine, mais son intention ne précise pas s'il faut le faire assis ou debout. Un problème pour un plan est un problème d'une autre sorte, car il remet le projet en question globalement ; par exemple, un mal de gorge et une fièvre peuvent faire que le conférencier se demande s'il doit effectivement réaliser son intention. Les problèmes pour les plans peuvent être liés soit à l'apparition d'obstacles, soit à l'émergence d'alternatives préférables. Plus généralement, Bratman propose que les agents ayant des dispositions raisonnables font attention à deux sortes de conditions soulevant des problèmes pour leurs plans, à savoir aux différences entre le monde tel qu'il était prévu au moment de la planification et le monde tel qu'il s'est effectivement réalisé, et aux des changements dans les préférences et les intentions de l'agent. L'agent raisonnable a tendance à prendre note de ce genre de changements ; s'il n'en observe pas, il a tendance à ne pas

reconsidérer ses intentions. Bratman avance une deuxième idée générale au sujet de la nature des dispositions raisonnables. Lorsque l'enjeu d'un certain projet est grand et que l'agent a une bonne occasion de le reconsidérer, l'agent ayant une disposition raisonnable en profitera en rouvrant le dossier[11].

De telles remarques ont du sens, mais elles ne constituent pas une théorie des dispositions raisonnables que peuvent avoir des agents par rapport à la reconsidération réfléchie et irréfléchie des intentions. Il est facile de fournir des exemples de cas faisant exception à la première thèse de Bratman à ce sujet. L'enjeu peut bien être grand et l'on peut bien avoir la possibilité de reconsidérer son intention, mais cela n'implique pas toujours qu'il soit rationnel de le faire, car l'on peut perdre son temps dans une délibération inutile, et celle-ci peut avoir des conséquences secondaires néfastes, comme un manque de confiance en soi ou un style de pensée inquiet et indécis. Prenons le cas des habitudes que l'on peut avoir par rapport au moment de s'endormir chaque nuit ; chacun a sa petite politique personnelle dans ce domaine. Nous avons tendance, je pense, à croire qu'il y a quand même un enjeu assez important ici, car tout le monde veut éviter l'insomnie. Nous avons aussi le temps nécessaire à une reconsidération de nos politiques du sommeil. Mais il n'est pas du tout raisonnable d'avoir tendance à reconsidérer nos intentions habituelles dans ce domaine, et il est même dangereux de s'engager dans ce genre de pensée. Celui qui se demande « que dois-je faire pour pouvoir m'endormir ? » a déjà un problème. Donc les dispositions raisonnables ne sont pas toujours (ou seulement) celles qui lient une croyance que l'enjeu est grand à une tendance à repenser le projet.

Y a-t-il d'autres critères auxquels l'on peut se référer afin de déterminer quand il est raisonnable d'avoir tendance à reconsidérer un certain type d'intention ? Peut-être. Une théorie de la nature des dispositions raisonnables exigerait une solide typologie des genres d'intentions et d'actions. Ensuite il faudrait spécifier le degré de stabilité qui lui serait subjectivement approprié. Construire une telle théorie serait une tâche énorme, car nous n'avons aucune typologie de cette sorte. Même si nous en avions

11. Voir citation, p. 67.

une, notre travail ne serait pas fini : le succès des agents dans la réalisation de leurs intérêts à longue échéance ne dépend pas seulement d'une disposition interne qui se manifeste dans le comportement, mais aussi des rapports archicomplexes qu'il peut y avoir entre le comportement et les circonstances externes. Tous ces rapports devraient être pris en compte par une théorie conséquentialiste de la disposition raisonnable. Une telle théorie, c'est de la *Zukunftsmusik*, c'est-à-dire la musique de l'avenir.

L'on peut soulever des questions au sujet de l'autre remarque générale sur la nature des dispositions raisonnables. Bratman dit qu'une propension raisonnable fait qu'un agent ne reconsidère pas son intention si l'agent ne remarque pas un problème pour son plan. Cela me semble juste, mais ce n'est pas du tout là une contrainte triviale sur la cognition de l'agent. Voici pourquoi. Décider qu'il y a un problème pour son plan peut consister en une perception assez routinière d'un événement. Tel est le cas du conférencier qui se réveille le matin avec un terrible mal de gorge. Il sait tout de suite qu'il y a un problème pour ses projets de la journée. Mais tous les problèmes pour les plans ne sont pas de cette nature ; l'agent doit aussi faire des inférences assez complexes au sujet des rapports entre le monde actuel et ses prévisions antérieures. L'agent doit décider si un écart tel ou tel entre les deux représente un problème pour son plan ou simplement un problème posé par son plan, et trancher cette question n'est pas toujours chose facile. Au fond, tout problème pour un plan est une question de trouver un moyen en vue d'une fin que l'on s'est donnée ; chaque fois que l'agent n'arrive pas à trouver un tel moyen, il y a un problème pour son plan. Dans certains cas, savoir qu'il y a un problème pour son plan consiste à inférer que tous les moyens possibles ont été considérés, ce qui est loin d'être une perception routinière.

L'idée derrière les remarques que je viens de faire est que le concept d'une disposition raisonnable de non-reconsidération ou de reconsidération prend la place, dans l'analyse de Bratman, d'une collection non spécifiée de processus cognitifs extrêmement complexes, car beaucoup de perceptions et d'inférences ont lieu entre le moment où l'agent forme son intention et le moment t_3 où il peut la reconsidérer ou non. L'objection que je suis en train de soulever contre l'approche de Bratman consiste à

dire qu'une théorie de la rationalité de l'agent devrait prendre en compte ces processus cognitifs, non pas par le biais d'une évaluation « globale » de leurs conséquences à longue échéance, mais directement, en évaluant leur cohérence, leur rationalité interne et leur fonction dans la délibération étendue. Le genre d'évaluation que je viens d'évoquer s'applique déjà, dans la théorie de Bratman, à tous les autres nœuds dans l'arbre de décision, y compris les moments où l'agent réfléchit sur l'opportunité de reconsidérer son projet. Pourquoi, donc, ne pas employer la même stratégie d'évaluation dans les cas du type DE4 et DE5 ? Une telle approche exigerait la révision suivante d'un principe de la rationalité de l'agent :

Principe révisé de la rationalité de la (non-)reconsidération réfléchie et irréfléchie

Il est rationnel pour S de (ne pas) reconsidérer, d'une manière réfléchie ou irréfléchie, son intention antérieure si et seulement si la (non-)reconsidération est au moins aussi bien fondée dans les raisons de S que l'alternative.

L'idée derrière le changement de principe que je propose est qu'en principe, nous pouvons procéder à une évaluation directe des processus cognitifs à l'œuvre et dans une (non-)reconsidération réfléchie et dans une (non-)reconsidération irréfléchie. Nous avons raison de préférer l'évaluation directe à une évaluation indirecte et seulement dispositionnelle. Imaginez, par exemple, un agent qui conclut à tort qu'il vient de buter contre un obstacle infranchissable pour un de ses projets ; sans y réfléchir, il entame tout un processus de reconsidération, se demandant s'il lui faut complètement abandonner son projet, et cela en dépit d'investissements importants et irrécupérables. Imaginons aussi qu'en entamant la reconsidération, l'agent se trompe[12], car une pensée irrationnelle fait qu'il ne voit pas qu'il a un moyen de mener à bien son projet. Dans la théorie de la rationalité de l'agent que nous propose Bratman, la pensée irrationnelle de cet agent n'est pas pertinente pour notre évaluation ; c'est la disposition plus

12. Un excellent exemple littéraire est le personnage de Lazare Chanteau dans *La Joie de vivre* d'Émile Zola ; pour un commentaire, voir mon étude *Literature and Rationality : Ideas of Agency in Theory and Fiction*, Cambridge, Cambridge University Press, 1991, chapitre 5.

générale dont elle est une manifestation, peut-être pas du tout typique, d'ailleurs, qui l'est. Je pense, au contraire, que nous devons nous intéresser directement aux processus cognitifs qui sont effectivement à l'œuvre dans la délibération étendue. Après tout, nous avons déjà vu qu'une théorie générale des dispositions raisonnables, c'est de la *Zukunftsmusik*, ce qui n'est pas le cas des évaluations d'exemples spécifiques de la (non-)reconsidération (ir)réfléchie de l'agent. Dans certains cas, par exemple, on sera en mesure de savoir qu'une reconsidération irréfléchie est le fruit d'un processus cognitif subjectivement inapproprié, ce qui mettra en question la rationalité de l'agent. Le principe révisé implique que, dans un tel cas, la non-reconsidération est irrationnelle ; le principe de la rationalité non-délibérative implique ensuite que la séquence de délibération étendue est un cas d'irrationalité de l'agent. Il s'agit là d'un jugement qui correspond maintenant à l'évaluation externe de la rationalité de l'action.

L'on pourrait objecter à cette approche que souvent nous n'avons aucun moyen de connaître les faits psychologiques requis par une évaluation directe de la cognition de l'agent, surtout lorsqu'il s'agit d'inférences inconscientes. Mais cette objection d'ordre épistémologique se retourne comme un gant : souvent nous n'avons aucun moyen d'évaluer les conséquences à longue échéance d'une disposition que nous avons déjà du mal à décrire ! Le principe révisé a au moins l'avantage de centrer l'évaluation sur le cas particulier de délibération étendue ; de plus, il traite les (non-)reconsidérations réfléchies et irréfléchies d'une manière symétrique. Si l'on me fait l'objection que des processus cognitifs irréfléchis ne sont pas sous le contrôle de l'agent, et ne peuvent pas être évalués en termes de rationalité, je réponds que la même critique peut être faite à l'encontre du principe de Bratman, car l'agent ne choisit ni ses habitudes ni les conséquences qu'elles peuvent avoir à long terme. Qui plus est, c'est un fait que nous portons des jugements sur nos comportements irréfléchis et sur nos actes et pensées manqués, même quand nous croyons que le comportement en question est inconscient et involontaire. Je ne vois pas pourquoi de tels jugements ne devraient pas faire partie d'une théorie de la rationalité de l'agent. L'approche de Bratman repose sur une distinction entre « la délibération

explicite » et « d'autres sortes de processus psychologiques »[13]. Mais il n'est pas évident qu'une telle distinction soit valable au sein d'une psychologie fonctionaliste, et même si l'on peut la défendre, il ne s'ensuit pas que ce ne sont que les délibérations explicites qui peuvent faire l'objet d'une évaluation en termes de rationalité. Un autre argument en faveur de la révision que je propose est que celle-ci rend les principes plus justes, car elle élimine chaque délibération étendue qui ne résout pas le dilemme du projet. Les principes de Bratman « pardonnent » des cas particuliers d'irrationalité, alors que le principe révisé ne le fait pas et décrit ainsi le genre de délibération que nous cherchons.

Une dernière remarque. L'auteur d'un compte rendu du livre de Bratman s'est demandé à quoi une théorie de la rationalité de l'agent pourrait bien servir. Tout ce qu'il nous faut, dit J. David Velleman, ce sont les normes pertinentes à la formation d'une intention, et ensuite à sa reconsidération[14]. Il est vrai, je pense, qu'une théorie de la rationalité de l'agent ne peut pas remplacer ce genre de normes, elle en dépend directement. Mais il est aussi vrai que nous nous intéressons aussi à des unités de comportement qui émergent lorsqu'un agent manifeste une séquence de moments de raisonnements pratiques interreliés les uns aux autres. Chaque moment dans la série doit satisfaire à certaines normes locales, mais il est aussi possible de procéder à une évaluation globale de l'ensemble, car des combinaisons différentes de moments locaux de rationalité et d'irrationalité peuvent mériter des évaluations distinctes sur le plan global. Une approche purement « atomique » de la rationalité ne saisit pas ce genre de différences entre des séries de raisonnements. Nous nous intéressons à des évaluations globales de délibérations pratiques étendues parce que nous savons que notre rationalité ne se réduit pas à une suite de coupes synchroniques, et notre capacité à négocier le dilemme du projet nous est hautement précieuse. Ainsi la théorie

---

13. Ainsi, dans « The Stability of Intentions », Bratman parle de « *psychological mechanisms of salience, problem detection, and the like* », en opposant ces mécanismes à ce qu'il appelle « *explicit deliberation* », p. 4, 7.

14. J. David Velleman, *Philosophical Review*, 100, 1991, p. 277-84 ; pour d'autres comptes rendus, voir Alfred R. Mele, *Mind*, 97, 1988, p. 632-4 ; et Frederick Adams, *Ethics*, 100, 1989, p. 198-9.

correspond à des intuitions et à des pratiques importantes, ce qui est précisément ce qu'elle devrait faire.

Département d'anglais
Université McGill