

RAISONS INTERNES ET NORMATIVITÉ¹

Introduction

Toute philosophie morale doit être en mesure d'offrir, de façon plus ou moins explicite, une analyse conceptuelle des relations d'un agent avec ses raisons, motifs, intentions et actions. La question de la normativité des raisons d'agir revêt à ce titre une importance particulière. Une problématique centrale dans les débats sur cette question concerne le type de relation qui doit être établi entre les raisons d'agir normatives et le rôle qu'elles sont censées jouer dans la motivation à accomplir les actes qu'elles justifient. J'aborde dans ce texte un aspect de cette problématique en discutant la distinction introduite par Bernard Williams entre raisons internes et raisons externes. À cet effet, la discussion partira d'un type d'explication plus général à l'intérieur duquel s'inscrit la distinction de Williams (1^{re} partie). Cela me permettra d'exposer, dans un premier temps, une difficulté que rencontre ce type d'explication plus général (2^e partie), puis, dans un second temps, de montrer que même si la distinction proposée par Williams et l'amendement à ce type général d'explication qu'elle permet d'introduire répondent précisément au désir de parer cette difficulté (3^e partie), la révision ainsi apportée sous la forme d'une défense des raisons internes y est malgré tout soumise (4^e partie).

1. Je tiens à exprimer des remerciements particuliers à Sylvain Fattebert. La pertinence de ses conseils et l'à-propos de ses encouragements ont constitué une aide précieuse à la rédaction de ce texte.

L'internalisme au sujet des raisons d'agir

Raisons normatives et raisons motivantes

Toute analyse des raisons d'agir semble devoir prendre en compte la distinction qu'il est possible d'établir entre les raisons motivantes et les raisons normatives, c'est-à-dire entre les raisons pour lesquelles un agent a accompli une certaine action et les raisons pour lesquelles cette action est susceptible ou non d'être justifiée, que la justification soit de nature morale, prudentielle, rationnelle ou d'un autre type. Ces deux sortes de raisons semblent assurément avoir pour caractéristique commune d'être des conditions d'intelligibilité d'une action : le fait de citer les raisons que possède une personne d'agir d'une certaine façon s'avère souvent indispensable à la possibilité de rendre intelligible le fait que cette personne a agi de la sorte ; il semble de plus probable de supposer que le fait de citer la raison normative qu'avait un agent d'agir d'une certaine façon soit, dans certains cas, une manière incontournable de rendre son action intelligible et irréductible au fait de souligner la raison motivante que pouvait par ailleurs posséder cet agent. Il n'en demeure pas moins que l'existence d'une telle distinction entre raisons d'agir motivantes et normatives semble être corroborée par l'observation du fait que les raisons pour lesquelles un agent a accompli une action peuvent ne pas être les raisons pour lesquelles cette action est justifiée, soit parce que cette action n'est précisément pas justifiée, soit parce que, bien qu'elle le soit, les raisons pour lesquelles elle l'est ne sont pas les raisons pour lesquelles l'agent l'a accomplie². Cette distinction entre raisons normatives et raisons motivantes a par exemple été exploitée par Michael Smith³ dans sa tentative de conciliation entre une théorie d'inspiration humienne de la motivation et une forme de réalisme au sujet des valeurs. Selon Smith, en effet, les considérations qui militent en faveur d'une certaine conception anti-humienne de la réalité des raisons normatives ne légitiment pas nécessairement leur transfert en faveur d'une conception anti-humienne de la théorie de la motivation.

La possibilité d'une telle distinction nous amène à nous poser la question de la relation existant entre les raisons d'agir normatives et leur rôle dans la motivation à accomplir les actes qu'elles justifient.

2. Sur la distinction entre raisons normatives et raisons motivantes, cf. l'introduction de Raz 1978.

3. Smith 1994.

De nombreux philosophes, qu'ils aient ou non cherché à combattre la rigueur de cette distinction, ont souligné qu'en général les raisons doivent au moins posséder une potentialité explicative des actions. Une raison d'agir, qu'elle soit normative ou motivante, doit être quelque chose au nom de quoi quelqu'un doit pouvoir agir. Comme l'affirme Williams :

S'il existe des raisons d'agir, il doit être le cas que les gens agissent parfois pour ces raisons, et s'ils le font, leurs raisons d'agir doivent figurer dans une explication correcte de leur action⁴.

Selon cette manière de voir, toute caractérisation de ce que c'est, pour une raison d'agir, que constituer une raison normative d'accomplir une certaine action devra donc, pour être adéquate, rendre compte de la manière dont cette raison est également susceptible de figurer dans une explication correcte d'une telle action.

L'internalisme au sujet des raisons d'agir

Même si le terme d'internalisme tend à devenir l'un des plus exploités et, partant, des plus intriqués et équivoques de la philosophie contemporaine, je propose, pour respecter un usage déjà bien établi⁵, d'appeler « internalisme au sujet des raisons d'agir » la position à laquelle je viens de faire allusion et selon laquelle toutes les raisons d'agir normatives doivent être des raisons d'agir motivantes possibles. Il est important de bien voir qu'il s'agit là d'une position que sont susceptibles de défendre *aussi bien* le partisan des raisons internes que le partisan des raisons externes au sens de la distinction de Williams. L'enjeu de la controverse autour de cette dernière distinction vise précisément à déterminer quelles sont les *implications* de l'internalisme au sujet des raisons d'agir sur l'interprétation qu'il convient de donner aux énoncés attribuant des raisons d'agir aux agents. Selon Williams, la conséquence d'une telle analyse est que ces énoncés doivent nécessairement être interprétés de manière *interne*, c'est-à-dire comme attribuant des raisons qui sont *relatives* aux motivations antécédentes de l'agent. Selon ses adversaires, au contraire, les raisons d'agir normatives peuvent parfaitement être reconnues comme des raisons motivantes possibles, même si elles sont des raisons attribuées au moyen d'énoncés recevant

4. Williams 1979, 102.

5. Johnson 1999, Brady 2000, Van Roojen 2000, Wiland 2000, par exemple.

seulement une interprétation *externe*, c'est-à-dire des raisons qui *ne* sont *pas* nécessairement relatives aux motivations antécédentes des agents auxquels elles sont attribuées.

Selon l'internalisme au sujet des raisons d'agir, la raison normative que possède un agent d'accomplir une certaine action dans certaines circonstances doit donc également pouvoir être la raison motivante pour laquelle l'agent accomplit cette action. Cette position impose par conséquent comme contrainte à toute analyse adéquate des raisons d'agir la capacité à faire comprendre comment la raison normative qu'il y a pour un agent A de ϕ (où ϕ désigne un verbe d'action) dans des circonstances C puisse également être la raison motivante qu'il y a pour A de ϕ dans C . On ne peut dire d'une raison R qu'elle est pour A une *raison* de ϕ dans C que si R peut figurer à la fois dans une explication intentionnelle de l'action de ϕ par A dans C et dans une justification rationnelle de cette action. Le défi qu'il s'agit de relever est de parvenir à concilier la dimension explicative et la dimension normative des raisons d'agir.

Une forme évidemment erronée d'internalisme au sujet des raisons d'agir consisterait à vouloir relever ce défi en exprimant la connexion entre les raisons et la motivation en termes d'une relation des raisons d'un agent avec ses désirs actuels. Une telle analyse prendrait la forme suivante :

(IR₁) Il y a une raison pour A de ϕ dans $C \supset A$ veut ϕ dans C .

Si cette thèse parvient assurément à rendre compte de l'explication de l'action de A , c'est au prix d'un sacrifice manifeste de sa connexion avec la justification de cette action. La solution généralement proposée par l'internalisme au sujet des raisons d'agir au défi qu'il lui incombe de relever consiste, par conséquent, à corriger cette analyse en affirmant l'existence d'une connexion qui serait *conditionnelle* entre ce qu'un agent a une raison de faire et ce qu'il serait motivé à faire *s'il était dans des conditions optimales*, ces dernières pouvant être spécifiées de différentes manières. Pour Michael Smith, par exemple :

dire que nous avons une raison normative de ϕ dans certaines circonstances C revient à dire que nous voudrions nous-mêmes ϕ dans C si nous étions pleinement rationnels⁶.

6. Smith 1994, 182. Cf. Korsgaard 1986, 321.

L'analyse internaliste doit dès lors prendre une forme conditionnelle du type⁷ :

(IR₂) Il y a une raison pour A de ϕ dans $C \supset$ Si A était pleinement rationnel, alors A voudrait ϕ dans C .

On peut également exprimer cette formule de la manière suivante :

(IR₂) Il y a une raison pour A de ϕ dans $C \supset A+$ voudrait ϕ dans C ,

où $A+$ représente A dans l'hypothèse où celui-ci est pleinement rationnel⁸.

La connexion entre le fait d'avoir une raison normative de ϕ et le fait d'être motivé à ϕ ne s'applique donc qu'à la condition que l'agent soit rationnel. Une telle théorie peut être comprise de manière caractéristique comme modelant les raisons d'un agent sur l'exemple de ce qu'une personne dotée de rationalité pratique serait motivée à faire. On peut par exemple envisager que $A+$ dans IR₂ réfère au *phronimos* aristotélicien. IR₂ pourrait alors exprimer la connexion entre le fait pour A d'avoir une raison de ϕ dans les circonstances C et le fait que l'action de ϕ est ce que le *phronimos* serait motivé à faire dans de telles circonstances. Ce modèle, que Michael Smith⁹ désigne par l'appellation du « modèle de l'exemple » par opposition au « modèle du conseil », peut évidemment également recevoir des caractérisations d'inspiration kantienne¹⁰.

Une difficulté pour l'internalisme au sujet des raisons d'agir

Il est possible d'énoncer trois conditions nécessaires auxquelles l'analyse IR₂ devrait au moins être en mesure de satisfaire afin de constituer une analyse acceptable des énoncés attribuant des raisons

7. Pour toutes les formalisations des analyses de l'internalisme au sujet des raisons d'agir, j'utiliserai le foncteur de l'implication matérielle, supposant ainsi que « le fait que A voudrait ϕ dans C si A était pleinement rationnel » représente une condition *nécessaire* au fait pour A d'avoir une raison de ϕ dans C . Certaines analyses – celle de Michael Smith ainsi que, même s'il semble avoir évolué sur ce point, celle de Williams lui-même – prennent la forme plus exigeante d'analyses en termes de conditions nécessaires et suffisantes. Je ne pense toutefois pas que l'argumentation soit affectée par ce point.

8. Cette formulation est empruntée à Railton 1986.

9. Smith 1995.

10. Par exemple Korsgaard 1986, 324.

d'agir aux agents. Étant donné que l'internalisme a pour ambition de parvenir à établir un lien entre les raisons normatives et les raisons motivantes, il devra tout d'abord respecter deux conditions faisant respectivement écho à la dimension explicative et à la dimension justificative des raisons d'agir. Il devra en premier lieu respecter ce que l'on pourrait appeler le « principe d'accessibilité à l'action » (P1), suivant lequel si R est une raison pour A de ϕ dans C, alors il doit être possible pour A de ϕ pour cette raison, et si A agit pour cette raison, celle-ci doit pouvoir figurer dans une explication correcte de son action. Une telle analyse devra en second lieu respecter le « principe d'accessibilité à la délibération » (P2), suivant lequel si R est une raison pour A de ϕ dans C, alors il doit être possible pour A d'en venir à reconnaître qu'il a cette raison d'agir au terme d'une délibération rationnelle. Enfin, et c'est le troisième principe (P3), l'analyse IR₂ devra bien évidemment être vraie. Je veux dire par là que l'équation IR₂ devra valoir pour tous les cas où l'énoncé « A a une raison normative de ϕ dans C » est vrai et ne devra pas autoriser que lui soient opposés des contre-exemples tels qu'un énoncé de ce type soit vrai bien qu'il soit faux que A voudrait ϕ dans C si A était pleinement rationnel. Pour le dire autrement, il devrait toujours être le cas, lorsque A a une raison normative de ϕ dans C,

- que si A était pleinement rationnel, alors A serait dans un état intentionnel qui pourrait *expliquer* le fait pour lui de ϕ ;
- que si A était pleinement rationnel, alors A serait *justifié* à ϕ .

Or, il est possible de mettre en évidence des situations où ces deux dernières conditions ne sont pas remplies. Robert Johnson a en effet montré que la position internaliste au sujet des raisons d'agir, dès lors qu'elle exprime ce type de connexion entre ce qu'un agent a une raison de faire et ce qu'il serait motivé à faire dans certaines conditions optimales, est sujette, comme toutes les conceptions philosophiques qui proposent des analyses conditionnelles, à ce que Robert Shope a appelé la « Conditional Fallacy » (CF)¹¹. Selon la définition qu'en donne Johnson, l'erreur caractéristique de ce type d'analyse

consiste en les différentes façons qu'il peut y avoir de négliger les dépendances pouvant exister entre l'*analysandum*, d'une part, l'antécédent et le conséquent du conditionnel dans l'*analysans*, d'autre

11. Cf. Shope 1978. Cette idée a notamment été reprise par Harman 1983.

part. On peut par exemple négliger un ensemble de contre-exemples à l'analyse tels que l'*analysandum* est assurément vrai, mais aussi tels que, l'antécédent du conditionnel dans l'*analysans* étant d'une manière ou d'une autre incompatible avec le conséquent, l'*analysans* est faux¹².

L'erreur intervient, pour le dire autrement, lorsqu'une définition est donnée en termes d'un contrefactuel et que la classe définie a des exemples qui impliquent la fausseté de l'antécédent du contrefactuel.

Reprenons l'analyse IR₂, qui établit une connexion entre le fait pour *A* d'avoir une raison de ϕ dans *C* et le fait que, *si l'agent était pleinement rationnel*, alors il serait motivé à accomplir cette action. On voit que IR₂ ainsi formulée est sujette à la CF, dans la mesure où l'idéalisation impliquée par le fait d'établir les conditions appropriées pour posséder une certaine raison d'agir et pour être motivé à agir d'une certaine manière peut s'avérer incompatible avec le fait pour un certain agent d'être dans les conditions qui fondent sa raison d'agir. Il est en effet possible pour un agent d'avoir une raison d'accomplir une certaine action bien que la pleine rationalité soit incompatible avec un quelconque désir d'accomplir cette action. C'est notamment le cas lorsque la raison que possède un agent d'agir d'une certaine manière est fondée sur sa condition d'agent non pleinement rationnel. Ainsi, le fait d'accomplir une action simplement dans le but de devenir davantage rationnel est un exemple de ce type. S'il est approprié pour un agent d'accomplir une telle action, alors la valeur de celle-ci dépend précisément du *manque* de rationalité de l'agent en question. Si cet agent était pleinement rationnel, il n'aurait aucune justification au fait d'accomplir une telle action et il n'aurait pas davantage de désir susceptible d'expliquer le fait de l'accomplir.

L'adoption de l'internalisme au sujet des raisons d'agir était censé garantir une connexion entre le fait pour un agent d'avoir une raison d'agir et le fait pour cet agent, s'il était pleinement rationnel, à la fois d'être dans un état intentionnel susceptible d'*expliquer* son action et d'être *justifié* à agir de la sorte. Or il existe manifestement des cas pour lesquels cette analyse est fautive puisqu'elle nous amène à négliger à la fois :

12. Johnson 1999, 54 ; voir également Millgram 1996, 211-212.

- que dans certains des cas pour lesquels il y a une raison pour A de ϕ , si A était pleinement rationnel, A ne serait pas dans un état intentionnel qui pourrait *expliquer* le fait pour lui de ϕ ;
- que dans certains des cas pour lesquels il y a une raison pour A de ϕ , si A était pleinement rationnel, A ne serait en rien *justifié* à ϕ .

Le problème de l’internalisme au sujet des raisons d’agir, si l’on en croit Johnson, c’est que toutes les tentatives offertes pour lui permettre d’échapper à ces contre-exemples et de se soustraire à la CF se soldent par un abandon de son attrait principal, à savoir celui de parvenir à concilier la dimension explicative et la dimension justificative des raisons d’agir. Aucune analyse ne serait, en d’autres termes, parvenue à satisfaire aux trois exigences exprimées par les principes énoncés précédemment. Ces exigences représentent pourtant autant de conditions nécessaires à la validité d’une telle analyse.

Or, il me semble que si la défense de la version amendée de l’internalisme au sujet des raisons d’agir que propose Williams en conséquence de l’élaboration de la distinction entre les raisons internes et les raisons externes répond précisément au désir de proposer une analyse qui soit soustraite à la CF, elle ne parvient pas davantage que les autres formes d’internalisme à satisfaire à ces trois exigences. Ce sont les deux étapes de cette affirmation que je vais successivement tenter de démontrer dans les deux sections qui suivent.

La distinction entre raisons internes et externes

Qu’est-ce qu’une raison d’agir interne ?

Dans une série d’articles célèbres¹³, Bernard Williams a proposé d’amender l’analyse IR₂ dans le sens de ce qu’il appelle une interprétation internaliste des énoncés attribuant des raisons d’agir aux agents. Un énoncé du type : « A a une raison de ϕ » doit selon lui être interprété de manière interne par opposition à une interprétation externe, étant entendu que l’énoncé « A a une raison de ϕ » reçoit une interprétation interne s’il est considéré comme signifiant

13. Williams 1979, 1989, 1995 et 1996. Une analyse des raisons d’agir proche de celle de Williams est également défendue par Harman 1975.

que « A pourrait arriver à la décision de ϕ au moyen d'une délibération valide conduite à partir de son S existant », où S est l'ensemble existant des désirs, préférences, évaluations et autres états psychologiques en vertu desquels il peut être motivé à agir¹⁴.

L'idée générale est qu'un agent a une raison de faire ce qu'un raisonnement pratique approprié, partant des divers éléments de son S , révélerait qu'il a une raison de faire, étant donné que :

- S , ce que Williams appelle « l'ensemble motivationnel subjectif » de l'agent, ne doit pas être compris uniquement en termes de désirs, mais également en termes des préférences, des dispositions évaluatives, des réponses émotionnelles, des loyautés personnelles et des projets divers de l'agent ;
- pour pouvoir conduire ce raisonnement pratique approprié et pour que sa délibération puisse être qualifiée de valide, l'agent doit satisfaire chacune des trois conditions suivantes :
 - il ne doit pas avoir de croyances fausses ;
 - il doit connaître toutes les croyances vraies pertinentes ;
 - il doit délibérer correctement, c'est-à-dire qu'il doit être à même de réfléchir sur les diverses façons dont la satisfaction des divers éléments de S peut être combinée, de découvrir de nouvelles possibilités, de déterminer la forme spécifique que prendra la réalisation d'un projet qui n'avait été adopté que de manière indéterminée, d'ordonner et de combiner dans le temps la satisfaction des différents éléments de S , de pondérer leur importance respective, d'en réviser l'intensité relative, la pertinence au contexte, l'urgence de leur satisfaction, etc.¹⁵.

On peut exprimer cette analyse au moyen de la formulation suivante :

(IR₃) A a une raison de ϕ dans $C \supset$ Si A était pleinement informé et délibérerait correctement à partir de son S existant, alors A pourrait arriver à la décision de ϕ dans C .

Ainsi, selon Williams, on ne peut dire d'un agent qu'il a une raison d'accomplir une certaine action si cette action ne permet pas par ailleurs de satisfaire un des éléments contingents de son

14. Cf., par exemple, Williams 1996, 66.

15. On peut comparer ces conditions de la délibération rationnelle avec celles énoncées par John Rawls 1971, § 63-64.

ensemble motivationnel subjectif. Cette conclusion peut être présentée comme la déduction de l'internalisme au sujet des raisons d'agir couplée avec la théorie humienne de la motivation. L'acceptation de l'internalisme au sujet des raisons d'agir implique en effet la reconnaissance de la validité de l'exigence motivationnelle suivant laquelle toutes les raisons d'agir normatives doivent être des raisons d'agir motivantes possibles. Quant à la théorie humienne de la motivation, elle suppose que le fait pour A d'accomplir intentionnellement l'action de ϕ implique que A possède au moins un élément contingent ψ de son ensemble motivationnel subjectif et une croyance que le fait pour lui de ϕ permet de satisfaire ψ . Ainsi, en couplant l'exigence motivationnelle stipulant que :

- (1) Toute raison d'agir normative pour A de ϕ doit être une raison d'agir motivante pour A de ϕ possible

avec la théorie humienne de la motivation exprimée par :

- (2) A a une raison motivante de ϕ dans C seulement si A a un élément contingent ψ de son S et une croyance que le fait pour A de ϕ permet de satisfaire ψ ,

Williams conclut que :

- (3) A a une raison normative de ϕ dans C seulement si A a un élément contingent ψ de son S tel que le fait pour A de ϕ permet de satisfaire ψ ¹⁶.

Ainsi, selon Williams, s'il y a un sens à parler de raisons normatives, celles-ci ne peuvent être au mieux qu'une sous-classe des raisons motivantes. En fait, les raisons d'agir normatives ne sont que ces raisons motivantes possibles pour lesquelles il est rationnel d'être motivé. L'inspiration humienne d'une telle conclusion trouve une illustration dans le fait que la théorie de la motivation s'étend au point de recouvrir l'ensemble de l'étendue occupée par la théorie des raisons pratiques. Il n'existe plus d'espace disponible pour d'autres types de raisons. Le point important que j'aimerais souligner, c'est que cette conclusion exprime toujours une forme d'internalisme au sujet des raisons d'agir, dans la mesure où elle admet l'existence d'une connexion conditionnelle entre ce qu'un

16. La référence à la croyance de A en la capacité de l'action de ϕ à satisfaire ψ est absente de (3), dans la mesure où, pour respecter la dimension normative de la raison d'agir, il doit être possible pour A d'ignorer le fait qu'il possède une raison normative de ϕ dans C .

agent a une raison de faire et ce qu'il serait motivé à faire s'il se trouvait dans certaines conditions idéalisées. Seulement, cette conclusion impose une restriction à la fois sur ce qu'un agent peut avoir une raison de faire et sur ce qu'il pourrait être motivé à faire, même dans certaines conditions idéalisées. Cette restriction est exprimée par l'idée d'une relativité aux éléments de l'ensemble motivationnel subjectif de l'agent.

Raisons internes et « Conditional Fallacy »

La défense des raisons internes proposée par Williams constitue donc une version amendée de l'internalisme au sujet des raisons d'agir. L'interprétation que j'entends soutenir est que le remplacement de IR₂ par IR₃ que cette défense appelle est motivé par l'idée selon laquelle seule cette dernière analyse serait capable de satisfaire aux trois conditions que doit nécessairement remplir toute forme acceptable d'internalisme au sujet des raisons d'agir. Cette interprétation, je le concède, n'est pas évidente à la lumière de la stratégie argumentative déployée par Williams dans son article « Internal and External Reasons ». La façon dont il y démontre l'incapacité de l'interprétation externaliste des énoncés attribuant des raisons d'agir – énoncée par IR₂ – de constituer une analyse satisfaisante de tels énoncés se concentre exclusivement sur les deux premières conditions, à savoir les deux principes d'accessibilité à l'action (P1) et d'accessibilité à la délibération (P2). L'argumentation, à ce stade, est indépendante d'une quelconque volonté de soustraire l'internalisme au sujet des raisons d'agir aux contre-exemples qu'il est possible de lui opposer. Elle suggère simplement que le remplacement de IR₂ par IR₃ est *déjà* justifié par l'incapacité dans laquelle se trouve IR₂ de satisfaire aux deux exigences exprimées par P1 et P2. Nous trouvons donc dans cet article l'élaboration de deux arguments contre la possibilité d'une interprétation externaliste des énoncés attribuant des raisons d'agir à un agent. Selon le premier argument – qui fait écho à la condition P1 –, une raison externe serait par elle-même insuffisante à l'explication d'une action qui serait choisie pour cette raison ; selon le deuxième argument – qui fait écho à la condition P2 –, même si l'on suppose que la conception des raisons externes parvient à satisfaire à l'exigence motivationnelle et explicative, elle n'apparaît pas en mesure de satisfaire à la deuxième exigence à laquelle doit répondre toute conception adéquate des raisons d'agir, à savoir que celles-ci doivent pouvoir être dégagées par l'agent au moyen d'une délibération rationnelle.

Les conditions P1 et P2 fournissent ainsi respectivement la prémisse des deux arguments de Williams contre la possibilité d'une interprétation externaliste des énoncés attribuant des raisons d'agir à un agent¹⁷.

Toutefois, même si le remplacement de IR₂ par IR₃ n'est pas motivé par le seul désir de soustraire l'analyse des raisons d'agir à la CF et même si la nécessité de satisfaire aux deux principes d'accessibilité à l'action et d'accessibilité à la délibération est censée justifier par elle-même un tel amendement à l'analyse, il n'en demeure pas moins que Williams affirme explicitement reconnaître dans l'incapacité de IR₂ de se soustraire à la CF, et donc de satisfaire au principe P3, une justification supplémentaire du bien-fondé de cette modification. C'est du moins l'argument qu'il développe dans sa réponse aux objections que lui oppose John McDowell¹⁸. Selon Williams, en effet, dans la mesure où elle ne cherche pas à faire dépendre les conclusions d'une délibération correcte d'un agent – et donc les raisons que peut avoir cet agent d'agir de telle ou telle manière – des éléments de son ensemble motivationnel existant, la perspective externaliste doit envisager que ce que cet agent a une raison de faire correspond à ce qu'il ferait s'il était un « délibérateur correct » *A+*. Supposons, même si c'est là l'objet d'une objection possible, que l'on puisse se mettre plus ou moins d'accord à la fois sur une conception de ce à quoi pourrait ressembler un « délibérateur correct » ou un *phronimos*, sur une conception de ce qu'un tel délibérateur correct serait motivé à faire dans les circonstances *C*, ainsi que sur l'idée que ce serait une bonne chose, dans l'ensemble, d'être un *phronimos*. Supposons, en d'autres termes, que nous ayons la connaissance d'une chose comme :

(G) Un délibérateur correct *A+*, un *phronimos*, serait motivé à ϕ dans les circonstances *C*.

Or, puisque, selon l'analyse IR₂, lorsqu'on dit de *A* qu'il a une raison de ϕ , on suppose qu'il y a un lien entre cette affirmation et l'idée que si *A* était un délibérateur correct *A+*, alors *A* serait motivé à ϕ dans les circonstances *C*, nous comprenons que IR₂, qui correspond à l'interprétation externe, reçoit en fait son contenu de G, au sens où IR₂ signifie seulement que G est appliqué à *A* et qu'il

17. On trouvera une reconstruction de ces deux arguments dans Millgram 1996, 198.

18. Cf. McDowell 1995 et Williams 1995.

est appliqué à *A* comme et parce qu'il peut être appliqué à tout agent, quelle que soit la composition particulière de son *S*. La conséquence est que, selon cette conception, l'analyse IR₂ ne permet pas d'envisager les énoncés attribuant une raison d'agir à un agent comme étant des énoncés portant de manière distinctive sur cet agent. Comme le dit Williams,

selon la conception externaliste ainsi construite, les énoncés [de ce] type ne relient pas des actions à des personnes, mais des types d'action à des types de circonstances, et ils sont exprimés de la manière la plus révélatrice sous la forme : « Dans les circonstances *C*, il y a une raison de ϕ »¹⁹.

Peu importe alors que la formule *G* soit exprimée ou non de manière formaliste et réglementée en termes de principes explicitement codifiés puisque, comme le montre l'exemple aristotélicien, même lorsqu'elle est énoncée en termes de ce que ferait une personne dotée de sagesse pratique sensible au contexte dans lequel elle se situe, ce qu'elle énonce n'en demeure pas moins impersonnel du point de vue des agents particuliers. Dans une perspective externaliste, la manière dont les énoncés invoquent la personne à laquelle ils attribuent une raison d'agir se limite à l'application d'une formule valable pour quiconque à un agent particulier, mais ne trouve pas son fondement dans ce qu'il y a de distinctif en cet agent particulier. Dès lors, cette conception traite « la personne *A* simplement comme un point de référence ou comme une occasion pour un jugement général »²⁰.

Or – et c'est ici qu'intervient le risque auquel s'expose une telle analyse de négliger les contre-exemples qui peuvent lui être apportés –, en considérant ce qu'il a une raison de faire, l'agent *A* doit prendre en compte dans quelle mesure, étant donné la composition particulière de son *S*, il n'est pas *A+*, par exemple dans quelle mesure il est dépourvu des vertus de caractère qui, de manière caractéristique, sont celles que possède le *phronimos*. Si l'agent *A* est conscient de ne pas faire preuve des vertus de caractère du *phronimos*, alors *A* peut, dans certaines circonstances, avoir de très bonnes raisons de ne pas faire ce qu'un homme doté de ces vertus accomplirait. C'est la raison pour laquelle IR₃ veille à ne pas faire dépendre ce qu'un agent *A* a une raison de faire dans certaines circonstances de ce que *A+* serait motivé à faire dans ces mêmes circonstances.

19. Williams 1995, 189-190.

20. *Ibid.*, 192.

On pourrait bien sûr rétorquer que l'analyse des raisons d'agir développée sous la forme de IR_2 est conditionnée aux circonstances particulières dans lesquelles se trouve l'agent. On pourrait donc faire la suggestion que les imperfections de l'agent, les traits de caractère qui le différencient de manière caractéristique de la figure du *phronimos*, doivent être considérées comme des modifications des *circonstances* de la délibération. Le problème de cette suggestion, c'est que

si les circonstances sont partiellement définies en termes des imperfections éthiques de l'agent, alors le *phronimos* ne peut se trouver dans *ces* circonstances, et G ne peut pas s'appliquer du tout²¹.

La seule solution raisonnable, selon Williams, si l'on veut considérer quelle est la raison d'agir de *A*, c'est sans nul doute de modifier l'interprétation de IR_2 afin de mieux prendre en compte, en acceptant de s'éloigner de G, la nature effective des dispositions de *A*. Mais « la question devient alors de savoir comment l'on peut faire cela sans, à terme, conduire à l'internalisme »²² au sens où Williams l'entend, c'est-à-dire à l'analyse IR_3 .

Une autre objection évidente qui pourrait être opposée à la conclusion de Williams consisterait à dire que si ce que nous poursuivons au moyen de l'analyse IR_2 , c'est une élucidation de l'énoncé « *A* a une raison de ϕ dans *C* », lorsque, par exemple, cet énoncé signifie que « *A* a une raison *morale* de ϕ dans *C* », alors le fait que *A* ne partage pas certaines des excellences de caractère de *A+* n'implique aucunement qu'il pourrait avoir une raison morale de faire quelque chose qui serait désapprouvé par *A+*. Il se peut que ce que *A* a une raison de faire ne soit pas ce que *A+* serait motivé à faire, mais la raison d'agir qui lui serait ainsi attribuée ne serait pas une raison morale, mais, par exemple, une raison prudentielle.

Il est néanmoins possible de répondre à cette objection de deux manières au moins. Premièrement, il serait sans doute erroné de

21. *Ibid.*, 190. Cf. Shope 1978, 412.

22. Williams 1995, 190-191. Une autre solution envisageable serait de prendre au sérieux l'idée selon laquelle le fait d'être un délibérateur correct représente une possibilité réelle pour tout agent. Le problème de cette solution, selon Williams, est qu'en plus des difficiles questions qu'elle doit affronter concernant le libre arbitre et la responsabilité qu'a un agent de son caractère, elle ne pourra certainement pas faire l'économie de l'idée que cette possibilité ne peut être qualifiée de réelle qu'à la lumière des développements et des expressions potentielles de l'ensemble motivationnel de l'agent, c'est-à-dire, finalement, des matériaux de l'internalisme (*ibid.*, 193).

considérer qu'il ne peut jamais être *moralement* significatif pour un agent *A* de tenir compte de ce qui le sépare de *A+*. Deuxièmement, Williams est prêt à admettre que l'énoncé « *A* a une raison de ϕ », interprété au sens de IR_2 , peut être asserté d'une manière pleinement intelligible. Seulement, lorsqu'il est asserté de cette façon, il a une signification qui pourrait aussi bien être exprimée au moyen d'un type de phrase différent, comme par exemple qu'il serait désirable que *A* fasse la chose en question. Une telle utilisation de cet énoncé implique néanmoins, comme le remarque Gilbert Harman, une équivoque entre le « devoir-être » (*ought to be*) et le « devoir-faire » (*ought to do*). Ainsi,

le fait qu'il serait merveilleux pour l'humanité si l'agent agissait d'une certaine façon n'établit pas en tant que tel que l'agent ait des raisons suffisantes d'agir de la sorte²³.

Si nous employons l'énoncé au sens de IR_2 pour blâmer le comportement de *A*, nous ne disons alors d'une certaine façon rien de plus que ce que nous cherchons à lui dire au moyen des différentes façons que nous pouvons avoir de lui reprocher d'être immoral. Si nous avons conscience de ce fait, alors il n'y a pas de problème. Seulement, le danger est que, lorsque l'on reproche à un agent de ne pas mettre son comportement en accord avec ce qu'il a une raison de faire, ce que nous suggérons par là, au sens distinctif du terme, c'est que l'agent est victime d'une déficience rationnelle ; nous l'accusons de ne pas avoir su délibérer correctement à partir des éléments de son *S*. Seule l'interprétation internaliste représente donc l'énoncé comme portant de manière *distinctive* sur les raisons de *A*. Si nous utilisons l'énoncé en question pour critiquer le comportement de *A*, et si nous l'employons dans le sens de l'interprétation internaliste, alors

la ligne de critique qui sera appropriée dans ces circonstances s'adressera à la performance de *A* en tant qu'agent rationnel ou raisonnable plutôt qu'aux autres déficiences qu'il peut avoir²⁴.

Critique de la position de Williams

L'analyse de Williams et la « Conditional Fallacy »

Nous avons vu que l'internalisme au sujet des raisons d'agir était censé garantir, sous la forme de l'analyse IR_2 , une connexion

23. Harman 1983, 161.

24. Williams 1996, 66.

entre le fait pour un agent d'avoir une raison d'agir et le fait pour cet agent, s'il était pleinement rationnel, à la fois d'être justifié à agir de manière correspondante et d'être dans un état intentionnel susceptible d'expliquer son action. Une telle analyse néglige pourtant les différentes façons dont l'idéalisation impliquée par le fait d'établir les conditions appropriées pour posséder une certaine raison d'agir peut s'avérer incompatible avec le fait, pour un certain agent, d'être dans les conditions qui fondent sa raison d'agir. Il est effectivement possible que, dans certaines circonstances, précisément celles dans lesquelles A n'est pas pleinement rationnel, A possède une raison de ϕ , quand bien même il ne serait en rien justifié à ϕ s'il était pleinement rationnel et quand bien même il lui serait, dans une pareille hypothèse, impossible de posséder le type d'état intentionnel susceptible d'expliquer une telle action.

L'analyse IR_3 , quant à elle, est supposée offrir une formulation de l'internalisme au sujet des raisons d'agir qui, du fait de la relativité à la composition caractéristique de l'ensemble motivationnel subjectif de l'agent qu'elle introduit, puisse se soustraire à des contre-exemples de ce type. Ce que A a une raison de faire n'est plus déterminé, selon cette analyse, par ce que son *Doppelgänger* pleinement rationnel serait motivé à faire, mais par ce que lui-même, étant donné la composition spécifique de son ensemble motivationnel subjectif, serait motivé à faire au terme d'une délibération rationnelle. Cette relativisation aux caractéristiques du S de l'agent semble éviter une dissociation entre les conditions qui fondent la raison d'agir d'un agent, d'une part, et les conditions appropriées à la reconnaissance de cette raison et à la motivation à agir de manière correspondante, d'autre part.

On peut toutefois se demander si l'analyse IR_3 , dès lors qu'elle affirme l'existence d'une connexion conditionnelle entre ce qu'un agent a une raison de faire et ce qu'il serait motivé à faire au terme d'un processus délibératif valide conduit à partir des différents éléments de son ensemble motivationnel subjectif, n'est pas également sujette à des contre-exemples semblables à ceux qui peuvent être opposés à IR_2 . De même que l'analyse IR_2 néglige le fait qu'il est possible pour un agent d'avoir une raison d'accomplir une certaine action qui ne serait pourtant ni explicable ni justifiable dans la perspective d'un agent pleinement rationnel, l'analyse IR_3 ne néglige-t-elle pas quant à elle le fait qu'un agent peut avoir une raison d'agir, relative à l'un au moins des éléments de son S , bien que l'élément même de son S qui la fonde soit incompatible avec la conduite d'une délibération valide ?

Imaginons, pour reprendre un exemple développé par Elijah Millgram²⁵, une théorie qui soit en mesure d'expliquer et de prédire nos aptitudes et nos inaptitudes cognitives et faisons la supposition que cette théorie est d'un niveau de complexité si élevé qu'il nous est impossible, précisément en raison de la nature de nos aptitudes et de nos inaptitudes cognitives, de la comprendre et, en conséquence, de l'apprécier. Une telle théorie ne peut être vraie que si nous sommes incapables d'en apprécier la vérité. Si nous parvenions à la comprendre et à en évaluer les mérites, alors elle serait par définition fautive, dans la mesure où les aptitudes cognitives que nous aurions dû déployer pour réussir à la comprendre auraient été celles-là mêmes dont la théorie expliquait pourquoi nous ne pouvions les posséder. De la même façon que la nature d'une telle théorie est conçue de sorte à rendre sa pertinence incompatible avec le fait d'en reconnaître la validité, un agent pourrait avoir une raison d'agir telle que l'élément qui en assure la justesse serait également celui qui rendrait irréalisable la conduite d'une délibération permettant d'en apprécier la validité. L'agent aurait alors une raison d'accomplir une certaine action qui ne serait ni explicable ni justifiable dans l'hypothèse où cet agent serait en mesure de conduire une délibération rationnelle valide.

Une autre manière de formuler l'équivalence offerte par Williams entre ce que *A* a une raison de faire et la décision à laquelle *A* pourrait arriver au moyen d'une délibération valide conduite à partir de son *S* existant consiste à affirmer que

ce que nous pouvons correctement attribuer [à un agent] au moyen d'un énoncé de raison interne à la troisième personne est également ce qu'il peut s'attribuer lui-même au terme d'une délibération²⁶.

Les contre-exemples que nous pourrions opposer à cette analyse seraient donc ceux qui démontreraient la possibilité qu'une raison d'agir soit correctement attribuée à un agent au moyen d'un énoncé interprété de façon interne, c'est-à-dire tenant compte des éléments de son *S*, bien que l'agent ne puisse pas arriver à la décision de *f* au moyen d'une délibération conduite à partir de son *S*. Ce qui nous autorise à attribuer correctement une raison d'agir à cet agent au moyen d'un énoncé à la troisième personne serait précisément ce qui lui interdit de se l'attribuer lui-même au terme d'une délibération conduite à partir des différents éléments de son *S*.

25. Millgram 1996, 201-202.

26. Williams 1979, 103.

Insensibilité et incapacité délibérative

Nous pouvons considérer le cas d'une personne faisant preuve d'insensibilité. Cette insensibilité peut être ce qui motive le fait pour cette personne de posséder une raison d'agir particulière. Mais cette insensibilité est également une incapacité délibérative : elle consiste, notamment, à être incapable d'apprécier la validité de certaines raisons d'agir. C'est même une caractéristique de l'insensibilité que de rendre celui qui en fait preuve incapable de reconnaître les raisons d'agir dont son insensibilité peut être la cause. S'il pouvait délibérer correctement et apprécier ces raisons d'agir, cela montrerait qu'il fait preuve de suffisamment de sensibilité pour *ne pas* avoir ces raisons d'agir. Comme le dit Millgram :

Dans un cas de ce genre, nous pouvons considérer qu'un agent a une raison de ϕ tout en rejetant l'affirmation suivant laquelle l'agent serait motivé à ϕ s'il délibérait correctement. Si, en effet, l'agent possédait la capacité rationnelle de délibérer de façon appropriée, alors il n'aurait pas la raison d'agir dont on affirme qu'elle se trouve au fondement de sa motivation²⁷.

Nous pouvons illustrer ce point à l'aide d'un exemple que Williams nous invite précisément à considérer, celui d'un agent *A* qui manifeste peu de considération et d'attention pour son épouse et qui reste insensible à toutes les injonctions ou à toutes les considérations que nous pouvons lui avancer afin de le convaincre qu'il a une raison de modifier son comportement. Cet agent, nous dit Williams, n'arrive pas à voir en quoi ces différentes considérations constitueraient une raison valable pour lui de corriger son attitude. Il n'y a, par hypothèse, aucun élément dans son ensemble subjectif motivationnel qui puisse lui donner une raison, au sens interne, de mieux traiter sa femme. Son ensemble motivationnel ne contient aucune disposition, même dissimulée et insoupçonnée, à répondre aux expériences et aux exercices de l'imagination susceptibles de lui montrer les bienfaits qu'il y aurait, même pour lui, à se comporter de manière différente : comme le souligne Williams,

Il n'y a *rien*, dans son ensemble motivationnel qui lui donne une raison d'être plus agréable avec sa femme dans l'état présent²⁸.

Considérons maintenant un agent *B*, semblable à *A* dans la mesure où il se comporte inlassablement de manière peu attentionnée

27. Millgram 1996, 203.

28. Williams 1989, 39.

avec sa femme. Disons que B est insensible et que ce trait de caractère fait partie de son ensemble motivationnel. Néanmoins, la femme de B , lassée par le comportement de son mari, s'est, à la différence de la femme de A , séparée de lui. À la suite de cette rupture, B désire ardemment se réconcilier avec sa femme. Et ce désir, également, est un élément de son ensemble subjectif motivationnel. Toutes les tentatives entreprises par B pour se rapprocher de sa femme depuis leur séparation se soldent néanmoins par des échecs, précisément parce que la façon dont il les entreprend témoigne de son manque de sensibilité et qu'elles ne font ainsi que conforter son épouse dans sa décision de rester séparée de lui.

Nous sommes confrontés ici à une situation où B possède, à la différence de A , une raison de changer son comportement. C'est ce que B a une raison de faire, au sens interne que Williams entend donner à ce genre d'énoncés, puisque c'est ce qu'il doit faire s'il entend satisfaire son désir de se réconcilier avec sa femme. C'est ce que nous pouvons correctement lui attribuer au moyen d'un énoncé de raison interne à la troisième personne. Mais est-ce également ce qu'il peut s'attribuer lui-même au terme d'une délibération conduite à partir des différents éléments de son S ? Il semble psychologiquement plausible de supposer qu'étant donné le manque de sensibilité de B , celui-ci, lorsqu'il délibère sur la meilleure façon de satisfaire son désir, ne parvient jamais à croire qu'il a une raison de changer son comportement, de faire preuve de davantage de sensibilité. B a donc une raison d'agir dont il est faux de dire que, s'il délibérait correctement, il serait motivé à agir de manière correspondante. Ce que B a une raison de faire semble, en contradiction avec P2, lui être inaccessible à la délibération et, si elle lui est inaccessible à la délibération, il est difficile de voir comment elle pourrait le motiver à agir. C'est de plus une raison telle que, si B possédait la capacité rationnelle de délibérer de façon appropriée, alors il ne serait pas dans les conditions qui en fondent la pertinence et la validité. Un énoncé de raisons internes correctement attribué à un agent à la troisième personne pourrait ainsi être vrai, bien que l'agent ne puisse pas arriver lui-même à atteindre la même conclusion au terme d'une délibération conduite à partir des éléments de son S .

Une objection possible

On pourrait m'objecter que ces remarques ne constituent pas un contre-exemple adressé à IR_3 comme l'étaient les contre-exemples opposés à IR_2 . Après tout, le cas de B est simplement le cas d'une

personne dont les éléments de son S sont en conflit les uns par rapport aux autres. Dans les contre-exemples adressés à IR_2 , l'analyse en question stipulait que ce que A a une raison de faire est ce qu'un $A+$, un A pleinement rationnel, serait motivé à faire. Les contre-exemples consistaient à mettre en évidence le fait que ce que A avait une raison de faire pouvait précisément venir du fait que A n'était pas $A+$, que A n'était pas pleinement rationnel. Dans le cas de IR_3 , ce qu'un agent a une raison de faire, c'est ce qu'il serait motivé à faire s'il était capable de mener une délibération valide à partir d'au moins un élément de son S . Il semble donc que pour qu'un contre-exemple similaire soit porté à charge de l'analyse, il faudrait que ce qui fonde la raison d'agir de l'agent soit précisément cela même qui fait qu'il n'est pas un bon délibérateur. Or, ce qui fonde la raison qu'a B de modifier son comportement, c'est simplement son désir de se réconcilier avec sa femme et non pas le fait qu'il est insensible, c'est-à-dire le fait qu'il est incapable de mener à bien une délibération valide. L'agent B a certes une raison de changer son comportement, c'est-à-dire de ne plus être insensible, mais cette raison, si c'est une raison interne, est relative à un élément de son S , et cet élément est le fait qu'il désire se réconcilier avec sa femme. Sa raison d'agir est relative à *cet* élément de son S et c'est bien parce que *cet* élément n'était pas présent chez A que celui-ci, bien que tout autant insensible que B , n'avait pas de raison, au sens interne, de changer son comportement.

Je pense toutefois que le problème reste entier pour l'analyse de Williams. Il est vrai que, d'un point de vue internaliste, le fait que B est insensible et qu'il serait désirable ou merveilleux qu'il ne le soit pas n'établit pas en tant que tel qu'il y ait des raisons suffisantes pour B d'agir de telle ou telle manière. B ne peut avoir une raison suffisante de modifier son comportement que dans la mesure où le fait de ne plus être insensible permet la satisfaction d'un des éléments de son ensemble motivationnel subjectif. Mais ce que cela montre, c'est tout simplement qu'en introduisant une relativisation des raisons d'agir d'un agent aux particularités des éléments de son ensemble motivationnel susceptible d'être satisfaits par l'action correspondante, l'interprétation IR_3 ne modifie en rien la résistance de l'internalisme au sujet des raisons d'agir à la CF. Peu importe que la justesse de l'attribution à un agent d'une raison de ne plus se comporter de manière déplaisante soit dépendante ou non de la présence au sein du S de cet agent d'un élément que cette action permet de satisfaire. Dès lors que l'analyse IR_3 partage avec IR_2 la caractéristique de faire dépendre ce qu'un agent a une raison

de faire de ce que cet agent serait motivé à faire s'il était dans certaines conditions idéalisées, elle s'expose au risque de négliger le fait que ce qu'un agent a une raison de faire peut précisément dépendre du fait qu'il ne se trouve pas dans ces conditions idéalisées. La question de savoir si les raisons d'agir d'un agent sont en outre relatives aux motivations antécédentes de cet agent est une question indépendante de celle de savoir comment éviter la CF. Bernard Williams a donc tort, me semble-t-il, d'associer ces deux problématiques et il n'apporte aucun éclaircissement véritable sur la façon dont l'analyse IR₃ serait, à l'inverse de l'analyse IR₂, en mesure de se soustraire à la CF.

Un dilemme pour l'analyse de Williams

J'ai déjà indiqué comment, selon Johnson, le problème de l'internalisme au sujet des raisons d'agir réside dans le fait que les diverses tentatives offertes pour lui permettre de se soustraire à la CF se soldent par un abandon de ce qui en constitue l'attrait principal, à savoir sa prétention à concilier la dimension explicative et la dimension normative des raisons d'agir. Les diverses tentatives de respecter le principe P3 ne paraissent pouvoir être menées à bien qu'au prix d'un sacrifice de l'un des deux autres principes P1 ou P2. Il me semble que le partisan de l'analyse de Williams est précisément confronté à un tel dilemme : deux stratégies s'offrent à lui, mais aucune ne parvient à concilier la dimension explicative avec la dimension normative des raisons d'agir.

Une première stratégie de repli que peut adopter le partisan de l'analyse IR₃ revient à accepter la conclusion de la section précédente et, partant, à nier le fait que, au sens distinctif que Williams veut donner aux énoncés de raisons d'agir, un agent puisse véritablement avoir une raison de ne plus être insensible, même si c'est là le seul moyen pour lui de satisfaire un désir existant. Une telle solution paraît toutefois peu satisfaisante. Elle semble impliquer qu'un agent comme *B* ne peut, du simple fait de son incapacité délibérative, avoir de raison de se comporter autrement qu'il ne le fait, même si ce serait la seule façon pour lui de satisfaire un désir qu'il possède par ailleurs. Cette solution de repli sacrifie ainsi manifestement la dimension normative et justificative des raisons d'agir : les raisons d'agir des agents ne sont plus seulement relatives aux caractéristiques de leur ensemble motivationnel, mais également aux limitations que ces mêmes caractéristiques sont susceptibles de faire peser sur leur capacité effective à mener une délibération valide.

La seconde stratégie qui s'offre au partisan de l'analyse de Williams consiste à reconnaître que l'impossibilité qu'il y a pour *B* de parvenir à la conclusion qu'il a une raison interne de ϕ dans *C* n'est qu'une impossibilité psychologique concrète. Or, pour garantir la dimension normative des énoncés attribuant des raisons d'agir à des agents, il ne faudrait pas faire dépendre ces énoncés des possibilités psychologiques concrètes des agents de parvenir à la décision de ϕ au moyen d'une délibération valide, mais des possibilités psychologiques *idéalisées et abstraites* de ces agents. C'est la solution qui semble recevoir la préférence de Williams lui-même, lequel affirme que l'idée d'une délibération valide dont dépend la conception internaliste des raisons d'agir doit bien être comprise sous « une forme idéalisée ou abstraite, qui écarte les effets psychologiques distinctifs du processus effectif de délibération »²⁹. Une telle conception de la délibération revient à écarter les effets psychologiques que les éléments de *S* peuvent avoir sur la capacité délibérative effective des agents en supposant que l'on peut tenir constant un élément du *S* de l'agent et se demander ce que l'agent ferait s'il était par ailleurs pleinement capable de conduire une délibération valide. Une telle interprétation de la position de Williams me paraît toutefois soumise à deux difficultés principales.

Elle amène en premier lieu à contester la pertinence pour le partisan des raisons internes d'établir une différence entre le cas de *A* et celui de *B*. L'analyse de Williams suppose que l'agent *B* possède une raison d'agir même si, en raison des défaillances qui affectent ses capacités délibératives, il peut se trouver dans l'impossibilité pratique de reconnaître la validité de cette raison. L'agent *A*, en revanche, bien que son ensemble motivationnel contienne les mêmes éléments défaillants que *B*, ne possède pas un désir particulier dont la satisfaction se trouve empêchée par les défaillances de son *S*. C'est l'absence de ce désir qui interdit de supposer que *A* possède une raison d'agir similaire à celle de *B*. Ce que Williams devrait toutefois démontrer, c'est dans quelle mesure la déficience qui affecte le processus délibératif de *A* n'est pas également ce qui l'empêche de reconnaître la valeur de certains états de choses qui, s'il en avait conscience, susciterait en lui un désir susceptible de lui conférer une raison interne de modifier son comportement. On voit mal, en effet, ce qui autorise Williams à postuler l'existence d'une frontière étanche isolant les compétences requises pour conduire un

29. Williams 1996, 67.

processus délibératif valide de la dimension cognitive indissociable des éléments conatifs dont est composé l'ensemble motivationnel subjectif d'un agent. Les préférences, les dispositions évaluatives et les réponses émotionnelles de l'agent qui façonnent un tel ensemble motivationnel constituent autant d'ingrédients déterminant la capacité à évaluer l'importance, la pertinence et la possibilité de satisfaction des différents éléments de S . Si la seule façon d'attribuer à B une raison de se comporter différemment consiste à supposer l'existence d'un processus de délibération qui implique un degré d'idéalisation au-delà des déficiences psychologiques particulières qu'il manifeste empiriquement, pourquoi ne pourrait-il pas en être de même dans le cas de A ? De manière plus générale, le fait de ne plus considérer la délibération comme un processus psychologique contingent ne peut conduire qu'à accentuer le caractère équivoque de l'affirmation suivant laquelle un agent a une raison de ϕ seulement s'il *devrait pouvoir* arriver à la décision de ϕ au moyen d'une délibération valide conduite à partir des éléments de son S . En effet, le fait de dire que l'agent devrait pouvoir y parvenir n'implique plus nécessairement qu'il soit *effectivement* en mesure de conduire la délibération en question :

un obstacle inconscient, comme le dit Williams, devra par exemple être levé avant que l'agent n'arrive à la motivation à ϕ ³⁰.

La difficulté à laquelle se heurte ici le partisan de l'internalisme réside dans la difficile distinction qui doit dès lors être introduite entre le fait, par exemple, de lever un obstacle inconscient afin de permettre l'expression de l'élément pertinent de l'ensemble motivationnel d'un agent et le fait d'affirmer qu'un tel élément est absent de son S et qu'il faudrait être capable de l'ajouter d'une manière ou d'une autre. Or, selon l'interprétation interne, il doit exister un critère de démarcation entre le fait d'en arriver, par la délibération à partir d'un certain ensemble motivationnel S existant, au projet de ϕ et le fait que la délibération donne une raison d'agir qui ne dépend pas de S . Il doit par exemple être possible de distinguer le cas où quelqu'un, conseillant un agent, aide ce dernier à découvrir ce qu'il a une raison de faire de la situation dans laquelle cette personne produit une différence dans l'ensemble motivationnel de l'agent et donne à ce dernier une raison d'agir qu'il n'avait pas auparavant. Si cette distinction se volatilise, c'est l'interprétation interne des

30. Williams 1995, 188.

énoncés attribuant des raisons d'agir qui, dans ce qu'elle a de distinctif par rapport à l'interprétation externe, se trouve menacée³¹.

Cette façon de concevoir la dépendance des raisons d'agir d'un agent par rapport à une conception idéalisée de ses capacités délibératives a en second lieu pour effet de rapprocher IR₃ de IR₂. Elle réintroduit en effet une différence entre l'agent auquel est attribuée une raison d'agir et l'agent qui serait motivé à agir de manière correspondante en pareilles circonstances. Ce que *A* a une raison de faire est à nouveau déterminé par ce que *A+* serait motivé à faire, où *A+* est conçu comme un agent idéal capable de conduire le processus permettant de déterminer ce qui constitue la pleine expression délibérative des éléments du *S* de *A*. Cette solution reviendrait en quelque sorte à considérer les imperfections de l'agent non plus comme des traits qui affectent la capacité délibérative elle-même, mais comme des modifications des circonstances de la délibération. Le problème de cette suggestion, comme je l'ai déjà indiqué, c'est que si ce qu'un agent pleinement rationnel est motivé à faire dépend certainement des circonstances dans lesquelles il se trouve, ces circonstances n'incluront toutefois jamais sa propre irrationalité.

Pour être en mesure d'éviter cette dernière difficulté, le partisan de l'internalisme au sujet des raisons d'agir n'aurait alors d'autre alternative que d'adopter le « modèle du conseil » préconisé par Michael Smith³² en lieu et place du « modèle de l'exemple » qu'exprime typiquement IR₂. Afin de saisir la différence entre ces deux modèles, Smith nous invite à imaginer deux mondes possibles : le monde *évalué*, dans lequel nous trouvons l'agent placé dans les circonstances auxquelles il est confronté, et le monde *d'évaluation*, dans lequel nous trouvons le moi pleinement rationnel de l'agent. Dans le modèle de l'exemple, la raison qu'il y a pour un agent de ϕ dans certaines circonstances du monde évalué dépend du fait que son moi pleinement rationnel placé dans le monde d'évaluation désirerait ϕ dans les circonstances du monde d'évaluation. On nous demande d'imaginer que le moi pleinement rationnel de l'agent met en œuvre son propre comportement dans son propre monde d'évaluation comme un *exemple* qui devrait être suivi par sa contrepartie située dans le monde évalué. Selon le modèle du conseil, au contraire, l'exigence internaliste est interprétée de telle manière que la raison qu'il y a pour un agent de ϕ dans les circonstances qui sont les siennes

31. Cf. Scanlon 1998, 369-370 et Cohon 1986, 555.

32. Smith 1995.

dans le monde évalué dépend du fait que son moi pleinement rationnel situé dans le monde d'évaluation désirerait qu'il accomplisse l'action de ϕ dans les circonstances du monde évalué. De cette façon, le modèle du conseil tient compte des cas dans lesquels ce qu'un agent a une raison de faire est en partie déterminé par le fait qu'il n'est pas pleinement rationnel. Un tel modèle peut être formulé de la manière suivante :

(IR₄) Il y a une raison pour A de ϕ dans $C \supset A+$ voudrait que $A\phi$ dans C .

Le problème de cette solution, c'est que c'est cette fois la dimension explicative des raisons d'agir qui est sacrifiée à sa dimension normative. Si l'interprétation IR₄ est en effet en mesure d'établir un lien qui peut sembler adéquat entre ce que A a une raison de faire et ce qu'il est justifié à faire, la dimension explicative de cette raison d'agir ne semble en revanche plus préservée dans la mesure où aucun lien n'est plus établi avec des états susceptibles de procurer une explication intentionnelle du fait pour A de ϕ . Les désirs auxquels cette interprétation rattache les raisons ne sont en effet pas des désirs *de faire* quoi que ce soit et ne sont donc en rien des motivations à agir. Les raisons d'agir, selon une telle analyse, sont ce que, si nous étions rationnels, nous désirerions *que* nous fassions dans des circonstances où nous ne sommes pas pleinement rationnels. Même si j'en viens à faire ce que mon double rationnel désire que je fasse, je ne serai néanmoins pas motivé par *ses* désirs que je fasse ces choses.

Conclusion

Ma compréhension de la défense par Williams de l'interprétation interne des énoncés attribuant des raisons d'agir aux agents est que cette interprétation doit être comprise comme une forme amendée de ce que j'ai appelé la position internaliste au sujet des raisons d'agir, à savoir la conception selon laquelle toute raison normative doit également être une raison motivante possible. Or, cette conception générale à l'intérieur de laquelle s'inscrit la position de Williams se heurte au problème de satisfaire aux trois conditions qu'elle doit pourtant nécessairement respecter afin de constituer une analyse acceptable des raisons d'agir. Il est notamment possible de mettre en évidence l'existence d'une difficulté qui ne semble pouvoir être contournée par l'internalisme au sujet des raisons

d'agir qu'au prix d'un abandon de ce qui en constitue l'attrait principal, à savoir sa prétention à concilier la dimension explicative et la dimension justificative des raisons d'agir.

L'amendement à la position internaliste réclamé par Williams est censé recevoir une double justification : non seulement la nouvelle interprétation de cette position serait la seule susceptible de satisfaire aux deux principes d'accessibilité à l'action et d'accessibilité à la délibération, l'internalisme dans son interprétation traditionnelle étant considéré comme incapable d'assurer la conciliation des deux dimensions correspondantes ; mais la correction suggérée serait encore la seule manière pour la position internaliste d'éluder la difficulté à laquelle se heurtent les autres interprétations qui en sont proposées.

La plupart des objections formulées à l'encontre de l'argumentation de Williams visent habituellement à contester la validité de la première forme de justification. Elles s'efforcent de démontrer qu'il n'est pas nécessaire de relativiser les raisons d'agir aux motivations antécédentes des agents auxquels elles sont attribuées pour rendre compte de leur dimension aussi bien explicative que normative. Indépendamment de la question de savoir si ces objections sont probantes ou non, je pense que la seconde justification que Williams entendait apporter à sa conception n'est pas recevable et que ce résultat dispense d'une certaine manière d'avoir à se prononcer sur la première forme de justification. Même si celle-ci s'avérait concluante, la variété d'internalisme qu'elle ratifierait n'en demeurerait pas moins invalidée par l'impossibilité dans laquelle elle se trouve de satisfaire simultanément à chacune des trois conditions précédemment énoncées.

Marc RÜEGGER

*Institut romand d'éthique
Université de Genève*

Références

- BRADY M.S. (2000), «How to Understand Internalism», *Philosophical Quarterly*, 50, p. 91-97.
- COHON R. (1986), «Are External Reasons Impossible?», *Ethics*, 96, p. 545-556.

- HARMAN G. (1975), «Moral Relativism Defended», *Philosophical Review*, 84, p. 3-22; repris dans *Explaining Value and Other Essays in Moral Philosophy*, Oxford, Clarendon Press, 2000, p. 3-19.
- HARMAN G. (1983), «Human Flourishing, Ethics, and Liberty», *Philosophy & Public Affairs*, 12, p. 307-322; repris dans *Explaining Value and Other Essays in Moral Philosophy*, Oxford, Clarendon Press, 2000, p. 151-164.
- JOHNSON R.N. (1999), «Internal Reasons and the Conditional Fallacy», *Philosophical Quarterly*, 49, p. 53-71.
- KORSGAARD C. (1986), «Scepticism about Practical Reason», *Journal of Philosophy*, 83, p. 5-25; repris dans *Creating the Kingdom of Ends*, Cambridge, Cambridge University Press, 1996, p. 311-334.
- MCDOWELL J. (1995), «Might There Be External Reasons?», in *World, Mind, and Ethics : Essays on the Ethical Philosophy of Bernard Williams*, J.E.J. ALTHAM & R. HARRISON (dir.), Cambridge, Cambridge University Press, p. 68-85.
- MILLGRAM E. (1996), «Williams' Argument Against External Reasons», *Noûs*, 30, p. 197-220.
- RAILTON P. (1986), «Moral Realism», *Philosophical Review*, 95, p. 163-207.
- RAWLS J. (1971), *A Theory of Justice*, Cambridge (Mass.), Harvard University Press; trad. fr. par C. Audard : *Théorie de la justice*, Paris, Seuil, 1987.
- RAZ J. (dir.) (1978), *Practical Reasoning*, Oxford, Oxford University Press.
- SCANLON T.M. (1998), *What We Owe to Each Other*, Cambridge (Mass.), Harvard University Press.
- SHOPE R.K. (1978), «The Conditional Fallacy in Contemporary Philosophy», *Journal of Philosophy*, 75, p. 397-413.
- SMITH M. (1994), *The Moral Problem*, Oxford, Blackwell.
- SMITH M. (1995), «Internal Reasons», *Philosophy & Phenomenological Research*, 55, p. 109-131.
- VAN ROOJEN M. (2000), «Motivational Internalism : A Somewhat Less Idealized Account», *Philosophical Quarterly*, 50, p. 233-241.
- WILAND E. (2000), «Good Advice and Rational Action», *Philosophy & Phenomenological Research*, 60, p. 561-569.
- WILLIAMS B. (1979), «Internal and External Reasons», in *Rational Action : Studies in Philosophy and Social Science*, R. HARRISON (dir.), Cambridge, Cambridge University Press, p. 17-28; repris dans *Moral Luck : Philosophical Papers 1973-1980*, Cambridge, Cambridge University Press, 1981, p. 101-113.

- WILLIAMS B. (1989), « Internal Reasons and the Obscurity of Blame », *Logos*, 10, p. 1-11 ; repris dans *Making Sense of Humanity, and Other Philosophical Papers 1982-1993*, Cambridge, Cambridge University Press, 1995, p. 35-45.
- WILLIAMS B. (1995), « Replies », in *World, Mind, and Ethics : Essays on the Ethical Philosophy of Bernard Williams*, J.E.J. ALTHAM & R. HARRISSON (dir.), Cambridge, Cambridge University Press, p. 185-224.
- WILLIAMS B. (1996), « Values, Reasons, and the Theory of Persuasion », in *Ethics, Rationality, and Economic Behaviour*, F. FARINA, F. HAHN & S. VANNUCCI (dir.), Oxford, Clarendon Press, p. 66-74.