Philonsorbonne

10 | 2016
Année 2015-2016

Irrationalité pratique et contrôle de soi par anticipation

Laurent JAFFRO

Édition électronique
URL : https://journals.openedition.org/philonsorbonne/810
DOI : 10.4000/philonsorbonne.810
ISSN : 2270-7336

Éditeur
Publications de la Sorbonne

Édition imprimée
Date de publication : 1 janvier 2016
Pagination : 131-152
ISSN : 1255-183X

Référence électronique
Laurent JAFFRO, « Irrationalité pratique et contrôle de soi par anticipation », *Philonsorbonne* [En ligne], 10 | 2016, mis en ligne le 20 janvier 2016, consulté le 08 juin 2021. URL : http://journals.openedition.org/philonsorbonne/810 ; DOI : https://doi.org/10.4000/philonsorbonne.810

© Tous droits réservés

Irrationalité pratique et contrôle de soi par anticipation

Laurent JAFFRO
(PHARE FRE3643-CNRS – Paris I)

Introduction

Prendre les hommes tels qu'ils sont et non tels qu'on voudrait qu'ils soient, voilà depuis Machiavel et Spinoza une maxime de la politique. Mais la faiblesse constitutive des agents humains, c'est-à-dire leur tendance à échouer, apparemment en l'absence d'obstacles, à adopter des conduites conformes aux valeurs sur lesquelles ils parviennent à s'accorder, est une bonne raison d'appliquer tout autant cette maxime à la morale.

Comme le formule Blaise Pascal, la « duplicité de l'homme est si visible qu'il y en a qui ont pensé que nous avons deux âmes, un sujet simple leur paraissant incapable de telles et si soudaines variétés, d'une présomption démesurée à un horrible abattement de cœur »[1]. Les « variétés » dont parle Pascal correspondent à des changements importants, voire à des inversions, de préférences. Ce sont des variations dans ce que je propose d'appeler l'identité *pratique*. Elles déstabilisent sérieusement les suites des ensembles de préférences d'un individu, sans affecter nécessairement son identité personnelle. Pascal indique ici que ces changements sont tellement radicaux que certains les interprètent comme des variations dans l'identité personnelle ou métaphysique (de telle sorte qu'il ne s'agirait plus du même sujet d'imputabilité). Dans la suite de ce propos, on considèrera seulement les cas où l'identité personnelle est préservée, de telle sorte qu'une véritable action sur soi, même si elle est indirecte, reste possible[2].

1. Blaise Pascal, *Pensées*, éd. L. Lafuma, 629.

2. Pour une illustration de la distinction entre identité personnelle et identité pratique, voir Laurent Jaffro, « Jekyll et Hyde. Quel modèle de multiplicité des soi ? », *Romanesques*, 7 (2015), p. 121-140.

Considérons maintenant le point de vue de Thomas Schelling, sur ce même sujet :

> Les gens se conduisent parfois comme s'ils avaient deux soi, un soi qui veut des poumons propres et une longue vie, et un autre soi qui adore le tabac ; un qui veut un corps svelte et un autre qui veut du dessert ; un qui est impatient de s'améliorer en lisant ce qu'Adam Smith dit de la maîtrise de soi <*self-command*> (dans *The Theory of Moral Sentiments*) et un autre qui préfère regarder un vieux film à la télévision. Ces deux soi sont continuellement en lutte pour le contrôle[3].

Il faut apprécier l'humour de la référence à Adam Smith, puisque la section sur le *self-command* dans la partie VI qu'Adam Smith ajoute dans la 6e édition de *La théorie des sentiments moraux* (1790) a pour sujet la maîtrise de soi à la stoïcienne, c'est-à-dire un contrôle direct et complet. Ce contrôle de haut niveau garantit la constance de nos résolutions à travers le temps ; il exclut la faiblesse de volonté (acrasie) ; plus généralement, il rend superflu tout recours à des techniques de contrôle indirect de soi. Si le *self* qui est en train de lire cette section de *La théorie des sentiments moraux* croit qu'il va trouver en elle une règle de vie qu'il pourra appliquer afin de prendre le contrôle du soi qui préfère regarder la télévision, il se trompe. Ce n'est pas en imitant le contrôle stoïcien des représentations que l'on peut y parvenir – ce serait une injonction impossible à mettre en œuvre – puisque, selon Thomas Schelling, le *problème* du contrôle se pose en partie parce que nous ne disposons pas du *self-command* que le stoïcisme nous attribue.

Selon Thomas Schelling, d'autres pourraient employer ce fait de la successivité des soi (que, pour sa part, il s'efforce simplement d'analyser) pour soulever des difficultés contre les fondements individualistes et utilitaristes de l'économie néoclassique[4]. En effet, la considération de ce fait suggère que dans certaines situations les agents cherchent à résoudre un problème de contrôle en plus d'un problème de décision : ils prétendent ne pas pouvoir se régler seulement et systématiquement sur leurs préférences actuelles, à n'importe quel moment de la période considérée, afin de les maximiser, car précisément ils veulent ne pas être un agent qui procédera ainsi dans le futur[5]. Ici j'emploie ce même fait pour soulever des difficultés contre une certaine espèce de rationalisme moral, de style kantien. Car le problème de *self-management* que pointe Schelling constitue une difficulté pour la philosophie de la décision au moins autant dans la tradition kantienne que dans la tradition utilitariste. Tandis que Thomas Schelling et Jon Elster ont tiré de ce type d'interrogations des enseignements relatifs à la philosophie économique, à la philosophie de l'action et à l'épistémologie

3. Thomas Schelling, *Choice and Consequence. Perspectives of an Errant Economist*. Cambridge, Mass., Harvard University Press, 1984, p. 59.

4. *Ibid.*

5. *Ibid*., p. 85.

des sciences sociales, l'objectif ici est de montrer son incidence dans un autre champ celui de la théorie morale.

Lorsque Schelling réfléchit aux situations dans lesquelles on a recours à une contrainte sur soi-même au moyen d'un tiers ou à un engagement public, parce que l'on sait que l'on risque sans cela de faire ou d'omettre de faire dans l'avenir ce que l'on préfère par dessus tout maintenant ne pas faire ou ne pas omettre de faire, le prototype de ces tactiques est fourni par Ulysse demandant à son équipage de se boucher les oreilles et de l'attacher au mât afin qu'il ne cède pas aux chants des sirènes, et même de lui désobéir s'il manifeste alors le souhait d'en être détaché (Elster a analysé en profondeur ce cas paradigmatique[6]). La morale ne peut intervenir, aux yeux de Schelling, dans cette affaire que sous une condition particulière : si l'on se pose la question de savoir si l'équipage a le devoir de répondre aux exigences de cet Ulysse présent qui programme les mesures de contrôle ou bien de cet Ulysse futur qui les subit. Cela correspond, par exemple, à des questions de déontologie médicale relatives à la fin de vie. Il est clair qu'en dehors des cas où le choix est facile – lorsqu'il ne fait pas de doute que la personne fait une demande extrêmement déraisonnable dans un état où elle est non seulement « inauthentique » (notion qui est cependant très relative et dépendante des représentations de soi), mais aussi manifestement hors d'elle-même parce que sous l'empire d'autre chose –, les cas les plus aigus pour l'éthique médicale concernent ceux dans lesquels les demandes de la personne au stade ultérieur qu'elle avait elle-même à l'avance disqualifiées (par exemple par un testament de vie) doivent être écoutées et entendues par le médecin[7].

Mais si l'on porte son attention vers la question de la moralité ou de l'immoralité de la conduite d'Ulysse lui-même, et non plus de l'équipage, on constate qu'elle ne se pose qu'à la condition qu'il soit moral ou immoral de céder aux chants des Sirènes. Si c'est immoral, alors il est moral de prendre des mesures préventives, par une dérivation de la valeur de la fin vers la valeur du moyen. Cependant, le rapport entre cette tactique et les valeurs morales est extrinsèque et contingent. Une éthique appuyée sur cette tactique est profondément hétéronome, là où la philosophie morale privilégie l'exigence d'autonomie. En effet, cette même tactique peut être employée dans, et justifiée par, un but immoral ou amoral. Il n'est pas du tout évident, d'ailleurs, que dans l'exemple homérique il soit question de morale. Pour cette raison, l'idée même de transposer dans le domaine éthique des considérations sur le contrôle de soi par anticipation pourrait paraître saugrenue.

Mais elle ne l'est pas. Pour comprendre la légitimité et l'intérêt de cette transposition de cette interrogation à l'éthique, il convient d'introduire

6. Jon Elster, *Le laboureur et ses enfants*, Paris, Minuit, 1987.

7. Sur ce genre de question, voir Schelling, *Choice and Consequence*, p. 88, 96 ; et sous une forme beaucoup plus développée, Joel Feinberg, *Harm to Self*, New York, Oxford University Press, 1986, p. 81-87.

une prémisse qui affecte la convergence, supposée par la plupart des philosophies morales, entre la moralité et la rationalité. Cette prémisse est celle de la faiblesse ou de la fragilité des agents[8]. Si nous nous considérons sous l'aspect de notre débilité constitutive – ce qui suppose que nous acceptions des vues très pessimistes de notre condition, qui sont semblables à celles de Blaise Pascal –, nous pouvons comprendre l'intérêt qu'il y aurait à substituer à une éthique pour super-héros une conception bien plus prudente.

La progression du propos est la suivante : sont d'abord présentées les conditions de la position d'un problème pratique relatif à la multiplicité des « soi ». Puis est discutée la nature rationnelle ou raisonnable des solutions pratiques que constituent les techniques de contrôle de soi par anticipation. Le problème et ses solutions invitent alors à réexaminer la place de l'irrationalité pratique dans la compréhension des liens entre jugement moral et motivation. Enfin, l'examen se porte sur quelques conséquences sur la conception de la liberté et de l'autonomie, et sur le rôle connexe des engagements dans la vie morale.

1. Comment entendre la successivité des soi ?

1. 1. Conditions nécessaires et suffisantes du problème de la successivité des soi

Les trois conditions suivantes, dont chacune n'est pas suffisante, constituent ensemble les conditions nécessaires et suffisantes pour que l'on puisse parler d'un problème de successivité des soi :

> A. Il y a des variations importantes (cette importance-là est une affaire de degrés), typiquement une inversion, du système des préférences importantes (cette importance-ci est relative à des considérations normatives) d'un même agent à travers le temps.
>
> B. Cet agent ne dispose pas de contrôle direct et complet de ses préférences durant la période concernée (pas de *self-command* au sens d'Adam Smith).
>
> C. Cet agent désire exercer un contrôle sur ses préférences futures.

La condition (A), à elle seule, ne suffit pas. En effet, on pourrait avoir la condition (A) remplie, mais avec un *superself*[9] qui assurerait la gestion des divers soi et la distribution entre eux. C'est ce qu'exclut la condition (B). Enfin, sans la condition (C), il n'y aurait pas de problème pratique pour l'agent concerné.

8. Certes, la prémisse de la fragilité s'applique mal au cas d'Ulysse, qui n'est certainement pas un agent faible par lui-même : sa faiblesse possible est entièrement relative à la force de séduction du chant des Sirènes.

9. Schelling, *Choice and Consequence*, p. 86.

Pour que ce problème puisse avoir une solution conforme aux vœux de l'agent, il faut que la condition (D) soit remplie :

D. Cet agent est capable de contrôle diachronique (indirect).

Dans le même esprit, on peut remarquer que la forme diachronique des cas d'irrationalité pratique, la succession de plusieurs soi au sens de l'identité pratique, ne diminue pas la capacité qu'a l'agent à assumer la responsabilité de l'ensemble des épisodes. C'est bien le même moi (au sens de l'identité personnelle) qui m'engage à ne pas céder à la tentation et qui y cède parce que plusieurs « soi » (au sens de l'identité pratique) sont en conflit. Comme le remarque Schelling, ces soi ont du mal à discuter entre eux et à négocier parce qu'ils ne sont pas simultanément réunis, alors que leur pluralité est bien interne à l'agent[10]. Cela éclaire la remarque de Pascal sur la « duplicité de l'homme ».

Il conviendrait, si l'espace de cette étude le permettait, de mieux distinguer le recours à une *successivité* des soi de la thèse d'une *multiplicité* des soi qui peut susciter un malentendu, celui d'une confusion avec des cas extrêmes de personnalités multiples ou avec un problème métaphysique de coexistence de plusieurs personnes en un corps unique. Parler de successivité, c'est marquer l'importance et la spécificité d'un problème de décision inter-temporelle[11]. Dans le cas du contrôle de soi par anticipation, ce problème est complexifié par l'instrumentalisation par l'agent des effets indirects de son action : le soi futur sera affecté de coûts ou de bénéfices qui ne seront pas l'effet de ses propres transactions, mais qui seront dus à l'action du soi passé ; le soi présent prévoit, quant à lui, d'être affecté dans le futur de ces coûts ou bénéfices, et choisit même d'en être affecté tout en sachant que, si cela ne dépendait que de son soi futur, ce dernier ne ferait pas un tel choix. Par conséquent ces coûts ou bénéfices constituent en quelque sorte des externalités pour le soi futur seulement, pour autant que son altérité peut être exagérée au point d'être considérée comme celle d'une autre personne ; pour le soi présent, ce sont des instruments d'action indirecte sur soi.

S'il faut tant insister sur cette distinction entre identité pratique et identité personnelle, c'est que la question « forensique » (comme disait Locke) de l'identité personnelle (c'est la question de la base métaphysique de la responsabilité morale au sens très large) est l'objet dans la tradition kantienne d'une sorte de surinvestissement normatif. Un bon exemple de cette moralisation de l'identité personnelle, sous le nom d'intégrité, se trouve dans les travaux de Christine Korsgaard, qui illustre à la suite de John Rawls l'école kantienne de Harvard. Cette auteure tend à confondre identité pratique et identité personnelle, puisque, selon elle, l'agent responsable de son action est l'agent dont les états psychologiques sont pleinement unifiés par la délibération :

10. *Ibid.*, p. 93-94.

11. Ce développement est redevable à des suggestions d'André Lapidus.

> L'œuvre de la délibération pratique est de réunification, de reconstitution : et la fonction des principes qui gouvernent la délibération – les principes de la raison pratique – est d'unification du soi[12].

Selon cette conception, l'action n'est vraiment accessible qu'à des agents fortement intégrés, habiles à se ressaisir eux-mêmes comme un centre non seulement unique, mais unitaire, d'action : « Pour qu'un mouvement soit mon action, pour qu'il soit une expression de moi-même à la manière dont une action doit l'être, il doit résulter de ma nature entière œuvrant comme un tout intégré[13] ». C'est une conception vraiment contraire à celle-là qui est défendue dans les sections qui suivent.

1. 2. Quelques caractéristiques des techniques de contrôle par anticipation

Il y aurait beaucoup de choses à dire sur les subtilités et la variété de ces techniques. Schelling et d'autres auteurs, notamment Elster, ont proposé des listes et même des typologies des techniques de ce contrôle diachronique. La liste suivante est empruntée à Schelling[14] : donner le contrôle à un tiers ; prendre un engagement, recourir à un contrat ; s'incapaciter soi-même ; modifier le contexte extérieur (par exemple, éviter les occasions de se mettre en colère) ; s'exposer à la surveillance ; s'emprisonner soi-même ; prévoir des récompenses ou des pénalités ; reprogrammer la série des actions (comme ne pas faire du shopping quand on vient d'être payé) ; s'appuyer sur les précurseurs ou déclencheurs (à rapprocher d'une recommandation de Jonathan Swift dans *For the Advancement of Religion* : il faut interdire le tabagisme à l'université parce qu'il est « parent du vice » de l'alcoolisme) ; temporiser, prévoir des délais (on peut penser au délai de rétractation pour certains achats) ; s'appuyer sur les autres, faire équipe ; automatiser la conduite, c'est-à-dire mettre en place un mécanisme qui agira malgré soi, etc.

Il faut prêter attention au fait que certaines de ces techniques de contrôle par anticipation (pas toutes) ont aussi une fonction cognitive, même si elle est secondaire. Si j'ai un recours à un tiers contrôleur, et s'il s'agit d'un agent capable de faire autre chose que de me contraindre, je peux aussi bénéficier de son avis. De manière plus générale, les techniques de contrôle indirect dont parlent Schelling et Elster ont leur équivalent du côté des techniques de limitation de l'erreur qui comportent fréquemment un appui sur un tiers, nos conduites épistémiques étant aussi des conduites sociales. Ce parallèle mériterait d'être exploré pour lui-même.

---

12. Christine Korsgaard, *Self-Constitution. Agency, Identity, and Integrity*, Oxford, Oxford University Press, 2009, p. 126.

13. *Ibid.*, p. 19.

14. Thomas Schelling, « Self-Command in Practice, in Policy, and in a Theory of Rational Choice », *The American Economic Review*, 74, 2 (1984), p. 1-11, p. 6-7.

Également, on peut remarquer que certaines tactiques du contrôle indirect consistent en une sorte de demande de tutelle paternaliste. Cet aspect de la discussion peut complexifier les débats sur l'acceptabilité morale du paternalisme. Généralement, le libéralisme considère que le *harm principle* de John Stuart Mill définit assez clairement les limites entre une intervention acceptable et une intervention inacceptable dans la vie d'autrui contre son gré. On voit bien cependant qu'un contrat d'Ulysse constitue de ce point de vue un cas énigmatique, puisque c'est à la demande du soi antérieur que la mesure est imposée par un tiers dans le futur, mais probablement contre le gré du soi postérieur. On peut dire que le soi antérieur a mis en place dans un esprit de paternalisme *soft* (puisque c'est lui qui en prend l'initiative) une mesure de contrainte par un tiers que le soi postérieur pourrait rejeter comme étant du paternalisme *hard* (contre sa volonté, et même contre son bien pour autant que le bien du soi futur est appréciable et définissable par lui-même).

Peut-on s'inspirer des stratégies interpersonnelles pour régler ces difficultés de gestion intrapersonnelle ? Peut-on voir dans ces difficultés un problème de coopération avec soi-même ? On peut s'appuyer sur la définition très générale de la coopération par Bernard Williams :

> Deux agents coopèrent quand ils s'engagent dans une entreprise commune dont l'issue dépend nécessairement des actions de chacun d'eux et dans un contexte où une action nécessaire d'au moins l'un d'entre eux n'est pas sous le contrôle immédiat de l'autre[15].

Selon B. Williams, la relation symétrique de coopération comporte ainsi une relation asymétrique de dépendance. Il y a au moins un « parti dépendant ».

Il me semble que, de même que je ne sais pas ce qu'autrui va faire dans le futur, et que je sais qu'il est bien possible et même probable qu'il ne fasse pas ce que j'attends qu'il fasse, de même je ne sais pas toujours ce que je ferai moi-même dans le futur, et il est possible et même probable que je ne fasse pas ce que j'attends que je fasse. Pour cette raison, mon action future n'est pas toujours sous mon contrôle immédiat, alors qu'elle fait partie des conditions nécessaires dont dépend le succès de mon projet. Si l'on suit la définition de Williams, le fait que le parti dépendant soit le *self* présent à l'égard du *self* futur n'empêche pas l'analogie avec un problème de coopération.

Cependant, dans le cas d'une coopération avec soi-même, la solution du problème semble facile à première vue, puisque je me connais assez moi-même, je suis assez familier de mes propres dispositions, de mes propres travers individuels, pour m'engager sans crainte dans cette coopération, selon une sorte de « jeu de l'assurance ». Mais, précisément, les choses se compliquent si, parmi mes dispositions qui sont bien connues de moi, il y a

15. Bernard Williams, « Formal Structures and Social Reality », in D. Gambetta (dir.), *Trust. Making and Breaking Cooperative Relations*, Oxford, Blackwell, 1988, p. 3-13, p. 7.

l'incapacité profonde à tenir mes engagements, et mon irrationalité pratique chronique. Dans ce cas, je me connais assez pour savoir que je ne suis pas digne de ma propre confiance. Dans cette situation de connaissance de mon propre manque de fiabilité, j'ai tout à craindre d'une telle tentative de coopération avec mon moi futur. Je ne suis pas vis-à-vis de moi-même dans la situation des deux rameurs de David Hume qui harmonisent leurs mouvements pour avancer sur l'eau (*Treatise of Human Nature*, III, II, 2), car il s'agit de mon comportement futur, et non pas de mon comportement observable maintenant, si bien qu'il ne saurait être question d'un ajustement automatique avec soi-même. Je suis plutôt, pour reprendre un autre exemple de Hume (*Treatise of Human Nature*, III, II, 5), dans la situation des deux paysans qui peuvent s'engager dans une collaboration dans des moissons qui sont décalées dans le temps, et dont le premier attend que le second lui retourne ultérieurement le service qu'il lui rend maintenant. Comme cette situation se répète année après année, les conduites de l'un et de l'autre finissent par s'ajuster, notamment parce qu'ils se soucient de leur réputation, mais le démarrage de la coopération a pu être très difficile.

Malheureusement, dans le cas du contrôle diachronique de soi, la répétition et l'expérience ne m'aident pas à lever les résistances, puisqu'il s'agit de la répétition et de l'expérience de ma propre irrationalité pratique. Dans ce cas, une solution raisonnable peut être le recours, entre autres moyens disponibles, à un tiers contraignant. À défaut de répondre correctement à mes propres exigences, je me place dans la situation d'avoir à répondre aux exigences d'autrui. Il y a une analogie évidente avec la manière dont la coopération peut se mettre en place s'il existe des incitateurs publics. Sachant que je suis incapable de rester fidèle dans l'avenir à ma résolution présente, je demande à un tiers de s'engager maintenant à me contraindre plus tard à respecter ma résolution, non pas *même si* plus tard je rejette cette résolution, mais *parce que je sais que* plus tard je risque de la rejeter. Je délègue ainsi à un tiers le soin de veiller à ce que ma résolution présente devienne difficilement révocable.

Il a été relevé plus haut que ces techniques ne sont pas intrinsèquement morales. Mais sont-elles au moins rationnelles ?

2. Le recours au contrôle de soi par anticipation est-il rationnel, raisonnable ?

2. 1. Physionomie du problème sous l'hypothèse (économiste) de la rationalité dite instrumentale

Selon David Hume (*Treatise of Human Nature*, II, III, 3), une passion (le terme est à prendre en un sens très large qui inclut aussi les désirs) ne peut pas être rationnelle ou irrationnelle, ne peut pas être conforme à la raison ou contredire la raison, sauf si elle est environnée :

(a) d'une information inexacte sur les circonstances, qui va jusqu'à l'erreur d'appréciation factuelle,

ou

(b) du choix de moyens inadaptés à la fin poursuivie.

Dire que la seule rationalité pratique que Hume admet est la « rationalité instrumentale » convient aux exigences relatives à (b) plutôt qu'à (a), mais on peut conserver cette terminologie.

Évaluons le recours au contrôle de soi par anticipation dans cette hypothèse humienne. Un tel recours semble à première vue rationnel au sens de la rationalité instrumentale. En effet si la fin poursuivie peut et même doit être atteinte par un moyen tel que l'évitement, le recours à un tiers contraignant, l'engagement public, ou je ne sais quelle autre technique du contrôle indirect, alors ce moyen peut être dit rationnel. Mais à y regarder de plus près ce recours est aussi bien irrationnel (de sorte que la question de sa rationalité est indécidable). Car simultanément :

– Le système futur de préférences, que je veux éviter d'avoir, fixe une fin contraire, au regard de laquelle le moyen que j'adopte est irrationnel.

– Mon système présent, initial, de préférences rend pourtant rationnel ce recours.

Bref, dans cette hypothèse il semble qu'on ait affaire à un cas d'importation dans les comparaisons intrapersonnelles des difficultés des comparaisons d'utilité interpersonnelles, dont leur indécidabilité[16]. Si l'on interprète la successivité des systèmes de préférences comme une succession de personnes distinctes, cette importation va de soi. Mais cette interprétation doit être nuancée. La nécessité de distinguer entre variations d'identité pratique et changements d'identité personnelle conduit à y voir seulement une analogie.

2. 2. Physionomie du problème sous l'hypothèse (moraliste) de la convergence de la rationalité et de la moralité

Passons de l'homme économique à l'homme moral et contredisons Hume d'un autre point de vue, de style kantien, qui identifie moralité et rationalité. La présentation qui suit s'appuie sur la formulation d'une théorie « kantienne » des raisons normatives par Michael Smith, dont l'ouvrage *The Moral Problem* (1994) a contribué à fixer les termes des discussions actuelles en métaéthique au sujet des raisons morales, du jugement moral et de la motivation morale. De ce point de vue kantien, pour paraphraser négativement Hume, il est *déraisonnable* de préférer la destruction de l'univers à un petit bobo à mon doigt, parce que ce n'est pas le genre de préférence que des agents pleinement rationnels recommanderaient d'avoir.

16. Schelling, *Choice and Consequence*, p. 92.

TSN (théorie smithienne des raisons normatives) : les désirs qu'on devrait (pour des raisons morales) avoir sont les désirs qu'on aurait dans une situation idéale de rationalité. Si j'étais pleinement rationnel, je n'aurais pas de préférence déraisonnable, déviante au regard de mes croyances à propos de ce qu'il est juste de faire dans la circonstance.

On note ici un recours à une rationalité superlative, comme si rationalité et irrationalité pratiques, dans cette conception, étaient une affaire de tout ou rien alors qu'un recours à une rationalité comparative pourrait suffire s'il s'agissait du recours épistémique à un point de vue moins partial.

Dans la conception de Michael Smith, la raisonnabilité est une propriété de la conduite cognitive et volitive d'agents idéaux. Mais les agents réels sont concernés par cela en ce qu'ils peuvent avoir des croyances évaluatives à propos de ces agents idéaux.

Appliquons maintenant cette conception à l'évaluation de la rationalité du recours au contrôle de soi par anticipation. Un tel recours est non seulement superflu, mais il a aussi quelque chose de déplacé, dans le cas d'un agent qui ne souffre pas d'une divergence entre rationalité et moralité et dont les motivations actuelles peuvent être affectées favorablement par ses croyances évaluatives à propos de ce qu'il préférerait s'il était un agent rationnel idéal. Le recours au contrôle de soi par anticipation demeure une option seulement pour les agents qui souffrent d'irrationalité pratique sévère. Seul le vice, ou du moins la défaillance chronique sur le chemin de la vertu, a besoin de telles béquilles. Mais de nouvelles questions se posent alors : comment l'irrationalité pratique est-elle possible ? Qu'en est-il si l'on adopte la prémisse de faiblesse constitutive ?

3. L'irrationalité pratique est-elle une exception qui défait un lien interne entre jugement moral et motivation ou bien une condition d'arrière-plan ?

3. 1. La thèse internaliste dans sa version « kantienne »

Michael Smith est un tenant de l'internalisme modéré, c'est-à-dire estime qu'il est vrai que (P) : si un agent juge qu'il est juste pour lui de φ <*φ : accomplir telle action*> dans les circonstances C, alors il est de ce fait même motivé à φ, sauf s'il souffre d'irrationalité pratique[17].

Soutenir (P), c'est affirmer que le jugement moral a une portée pratique et défendre l'internalisme du jugement, c'est-à-dire la thèse d'un certain lien intrinsèque mais défaisable entre jugement moral et motivation.

La première partie du livre de Michael Smith est consacrée à la question suivante : si l'on soutient (P), quel est l'ingrédient dans le jugement qui est

17. Michael Smith, *The Moral Problem*, Oxford, Blackwell, 1994, p. 61.

responsable de la motivation ? La réponse à cette question consiste en une discussion de la « théorie humienne de la motivation » (THM), que Smith conserve et aménage, tandis qu'il rejette la « théorie humienne des raisons normatives » (THN). Les humiens (et David Hume lui-même) admettent la distinction *conceptuelle* entre raisons normatives et raisons motivantes, mais soutiennent la thèse *existentielle* selon laquelle les raisons normatives ne sont rien d'autre, en fait, que des raisons motivantes. Selon (THN), il y a coïncidence de fait des sources psychologiques de la motivation et de celles de l'évaluation.

À cela s'oppose la théorie smithienne des raisons normatives (TSN, ou « rationalisme moral ») :

> Dire qu'accomplir une action d'une certaine sorte en certaines circonstances est juste <…>, c'est dire entre autres choses qu'il existe une raison normative de l'accomplir. Et ceci équivaut simplement à dire que des créatures pleinement rationnelles auraient le désir qu'une telle action soit accomplie dans de telles circonstances, si un tel désir a le contenu approprié[18].

La croyance portant sur ce qu'un agent ferait dans une situation contrefactuelle (idéale) de rationalité rend compte de l'objectivité, de la non-relativité à l'agent, et de la publicité des raisons normatives. Comprendre ce que signifie « avoir une raison normative de φ », c'est croire que dans la situation où nous serions pleinement rationnels nous voudrions accomplir φ (nous serions motivés à φ).

Michael Smith soutient à la fois (P) et (TSN) : Si un agent juge qu'il est juste pour lui de φ dans les circonstances C parce qu'il croit que s'il était pleinement rationnel alors il voudrait φ etc., alors il est motivé à φ, sauf s'il souffre d'irrationalité pratique. Selon Smith, (TSN) rend bien compte de (P). En effet, dans la mesure où nous sommes plus ou moins rationnels,

– soit, sans avoir besoin d'un autre appui, nous désirons déjà φ qui est l'action qu'il est juste d'accomplir dans les circonstances C,

– soit cette croyance évaluative que φ est l'action qu'il est juste d'accomplir cause notre désir de φ,

– soit encore nous nous efforçons de mettre en cohérence notre croyance évaluative et nos désirs. Si nous ne désirons pas φ alors que nous avons cette croyance évaluative, c'est que nous ne sommes pas pleinement rationnels[19].

Sous la plume de Michael Smith, la faiblesse de volonté (acrasie) est constamment quelque chose dont un agent « souffre » et est mise ainsi sur le même pied que les dépressions et autres pathologies. Je soutiens, au contraire, une thèse externaliste à propos du lien entre jugement moral et motivation : la déconnexion entre jugement et motivation est une condition d'arrière-plan que les moralistes diffament sous divers noms

18. *Ibid.*, p. 200.

19. *Ibid.*, p. 193.

de pathologies. L'irrationalité pratique à un certain degré est le lot commun, et une bonne psychologie morale devrait partir de cela plutôt que de la supposition inverse[20].

3. 2. Contre la thèse internaliste dans sa version « kantienne »

Argumentons dialectiquement contre (TSN) :

Je peux savoir que j'ai une raison normative de φ et ne pas être motivé à φ. C'est un cas d'irrationalité pratique. Comment l'analyser ? J'ai la croyance évaluative qu'un agent pleinement rationnel désirerait φ. Je n'ai pas ce désir. Cela signifie simplement que la motivation qu'aurait un agent rationnel de φ n'est pas la mienne. Les raisons normatives n'ont pas d'autorité suffisante pour moi. Selon (TSN), dans cette situation, je sais que je manque de ce désir, et la conscience de ce manque, la croyance évaluative qui permet de le repérer, peut causer (de manière instable, dépendante de facteurs tels que le tempérament ou l'éducation) mon désir de φ[21].

Selon (TSN), ce qui me manque dans cette situation d'irrationalité pratique n'est pas la croyance qu'un agent pleinement rationnel voudrait φ ; me manque la chance que certaines conditions soient remplies (celles dans lesquelles je serais un agent pleinement rationnel, ou bien celles, différentes et plus compatibles avec la faiblesse humaine, dans lesquelles ma croyance évaluative causerait mon désir de φ), car ces conditions manquantes sont celles qui constituent le contexte dans lequel j'aurais le désir de φ. La thèse est alors, malgré elle, externaliste. Car (P) suppose que ce soit du fait même du jugement moral que je sois motivé à φ (lien conceptuel, intrinsèque). (THM/THN) garantit (P) car le lien entre jugement et motivation, assuré par les désirs, est l'arrière-plan de tout jugement moral. (TSN) fragilise (P) parce qu'elle soutient qu'en l'absence de désir de φ la croyance qu'un agent rationnel désirerait φ peut suffire à causer le désir de φ : c'est admettre que cette croyance peut ne pas entraîner ce désir et substituer un rapport causal à un lien conceptuel. Soit les agents sont, en général et en raison de l'absence de lien conceptuel entre jugement et motivation, à divers degrés pratiquement irrationnels dans la situation réelle si certaines conditions extérieures au jugement moral qui font la « connexion » ne sont pas remplies (thèse externaliste, à laquelle je souscris), soit l'irrationalité des agents est une anomalie qui vient défaire une connexion établie avec la motivation au sein du jugement moral (thèse internaliste), mais cela conforte (THN) contre le projet de Michael Smith.

Mettons en rapport cette critique de (TSN) et la question qui fait l'objet de cette étude. La croyance (que l'on peut dire « pessimiste ») que peut avoir un agent irrationnel, *qu'il ne désire pas comme il devrait désirer*, joue un

20. Sur cette appréciation des liens entre morale et psychologie morale, il faut donner raison à Michael Stocker, « Desiring the Bad: An Essay in Moral Psychology », *The Journal of Philosophy*, 76, 12 (1979), p. 738-753.

21. Michael Smith, *The Moral Problem*, p. 193.

rôle important dans la vie morale. Cette croyance est-elle identique à la croyance évaluative, dont le rôle est essentiel selon le « rationalisme moral », *selon laquelle un agent pleinement rationnel désirerait comme il devrait désirer* ? Non, car ce sont des croyances sous des descriptions différentes, dont les contenus sont différents ; mais ce sont des croyances très voisines, qui peuvent se combiner et se suggérer l'une l'autre, si on admet qu'on peut avoir cette croyance évaluative-ci et simultanément la croyance qu'il n'y a pas d'agent pleinement rationnel ou du moins pas dans ce monde-ci. Cependant, la croyance pessimiste nous conduit à un contrôle des désirs très différent de celui qu'envisage le rationalisme moral.

4. Interprétation du problème de la convergence entre moralité et rationalité en termes de liberté et de système hiérarchique de désirs et de volitions

4. 1. La liberté de la volonté accessible à des agents faibles est diachronique

La discordance entre évaluation (sensibilité aux raisons normatives) et motivation qui caractérise de manière générale la faiblesse de volonté et sans doute d'autres formes d'irrationalité pratique peut être décrite par la conjonction suivante :

Je ne veux pas comme je devrais vouloir
et simultanément
je voudrais vouloir comme je devrais vouloir.

Cette conjonction peut être lue de manière synchronique ou bien diachronique. L'existence d'une faiblesse de volonté synchronique (qu'envisage Donald Davidson et que Jon Elster dans ses travaux récents appelle acrasie « au sens strict », dont il doute qu'elle soit un fait avéré) n'est pas supposée par le présent propos. En revanche, un contrôle synchronique est exclu. S'il n'y a pas de faiblesse de volonté synchronique, ce n'est pas nécessairement parce que nous disposons d'un contrôle synchronique et direct.

Remarques préalables. Ici, « je voudrais » ne désigne pas une volition, mais un souhait. « Je veux » désigne une volition. « Je voudrais vouloir » et « je veux vouloir », pour cette raison, ne portent pas sur le même type d'opération. Il est indéniable que nous avons des souhaits à propos de nos volitions ; il est douteux que nos volitions puissent avoir directement pour objet d'autres volitions. Distinguons des volitions de 1er ordre (notées « vouloir$_{\mathbf{1}}$ ») et de 2nd ordre (notées « vouloir$_{\mathbf{2}}$ ») et observons que :

Si pour vouloir$_1$ comme je voudrais vouloir$_1$, il *suffisait* que je *veuille*$_2$ vouloir$_1$ comme je voudrais vouloir$_1$, c'est-à-dire si j'avais un contrôle *direct* et *synchronique* du vouloir$_1$ par le vouloir$_2$, alors :

(1) j'aurais le type et le degré (très élevé) de liberté que la philosophie a conçue sous le nom de libre-arbitre ;

(2) la liberté et la rationalité pratique s'impliqueraient réciproquement (sauf blocage, contrainte, ou manipulation externe, j'aurais toujours les préférences que je veux avoir) ;

(3) il n'y aurait plus de cas d'irrationalité pratique de ce type (j'aurais toujours les préférences que je voudrais et devrais avoir), sauf pour les êtres dont le libre-arbitre dysfonctionne ; bref, la rationalité pratique et une forme très élevée d'autonomie s'impliqueraient réciproquement.

Or nous ne disposons pas d'un tel contrôle direct et synchronique du vouloir$_1$ par un vouloir$_2$. Nous disposons d'un contrôle indirect et diachronique, comme l'ont bien vu plusieurs auteurs, dont Leibniz. Selon lui, nous sommes incapables de vouloir à volonté comme de croire ou juger à volonté[22] :

> Par des actions volontaires nous contribuons souvent indirectement à d'autres actions volontaires, et quoiqu'on ne puisse point vouloir ce qu'on veut, comme on ne peut pas même juger ce qu'on veut, on peut pourtant faire en sorte par avance qu'on juge ou veuille avec le temps ce qu'on souhaiterait de pouvoir vouloir ou juger aujourd'hui[23].

Notons qu'ici « ce qu'on veut » veut dire « comme on veut ». Également, il y a un parallèle parfait entre l'absence de contrôle direct de la volonté sur le jugement ou sur la croyance et l'absence de contrôle direct de la volonté sur elle-même. Dans la suite de ce passage, au chapitre 21 du livre II des *Nouveaux Essais*, qu'il faudrait commenter en détail, Leibniz a manifestement sous les yeux les passages de Blaise Pascal qui insistent sur la manière dont, bien que les croyances ne soient pas volontaires, nous pouvons finir indirectement (notamment grâce au pari) par avoir les croyances qui nous arrangent. Comme Pascal, Leibniz estime que si nous étions en présence du vrai bien – dans un genre de situation idéale semblable à celle que décrit le kantisme contemporain lorsqu'il parle des réponses des agents idéaux à des raisons morales – nous ne saurions alors manquer de l'aimer de tout notre cœur ; mais notre condition de faiblesse est telle, l'irrationalité pratique est si ordinaire, qu'il nous faut adopter une tactique de contrôle par anticipation :

22. Sur cette question classique, voir les études réunies dans L. Jaffro (dir.), *Croit-on comme on veut ? Histoire d'une controverse*, Paris, Vrin, 2013.

23. Leibniz, *Nouveaux Essais sur l'entendement humain*, éd. J. Brunschwig, Paris, Garnier-Flammarion, 1966, p. 155-156.

> Lorsqu'un homme est dans de bons mouvements, il doit se faire des lois et des règlements pour l'avenir, et les exécuter avec rigueur, s'arracher aux occasions capables de corrompre, ou brusquement ou peu à peu, selon la nature de la chose. Un voyage entrepris tout exprès guérira un amant ; une retraite nous tirera des compagnies qui entretiennent quelque mauvaise inclination[24].

Il est remarquable que Leibniz mentionne parmi ces techniques de contrôle indirect un ingrédient traditionnel des exercices spirituels, dont la philosophie française contemporaine, avec Michel Foucault et Pierre Hadot, a fait son miel. Mais Leibniz lui assigne à raison une fonction bien plus modeste que celle de la « subjectivation ». Leibniz suggère que la retraite n'est pas simplement la situation dans laquelle il convient de se placer si l'on vise à développer le *self-command* stoïcien, mais aussi et plus raisonnablement un des outils qui conviennent à ceux qui sont incapables d'un tel *self-command*. De manière plus générale, il conviendrait de réexaminer la question des stratégies de second choix (entendu relativement à un premier choix qui, sous l'hypothèse de la faiblesse, n'est jamais accessible) de promotion de la sagesse et être plus attentif à l'importance du contrôle par anticipation y compris chez les auteurs qui souscrivent à l'intellectualisme socratique sous ses formes les plus fortes[25].

4. 2. Contre une conception forte de l'intégrité

Le modèle de Leibniz s'oppose très fortement à celui des métapréférences lorsque celles-ci sont conçues comme étant *contemporaines* des préférences qu'elles permettent de contrôler, comme c'est le cas, il me semble, dans une des plus subtiles analyses contemporaines de la liberté psychologique, celle de Harry Frankfurt. Si, comme je le crois, nous est accessible en guise de contrôle de nos volitions seulement un contrôle diachronique, alors agir de tout son cœur (le *wholeheartedness*, dont Frankfurt fait en quelque sorte le sommet de l'agentivité (*agency*), et qui est une harmonie entre nos désirs de second ordre et nos désirs de premier ordre, ces derniers étant alors les désirs que nous désirons avoir et que nous faisons ainsi vraiment nôtres[26]) est quelque chose de très ambigu. C'est soit une liberté – le libre-arbitre – que nous vantent certains philosophes mais qui nous est inaccessible dans les faits, soit, plus raisonnablement, quelque chose qui marque une sorte de confort intérieur, mais qui peut aussi être l'indice, il me semble, d'un manque de contrôle de soi. En effet, l'acratique, au moment où, par exemple, il s'adonne au plaisir qu'il désapprouve, peut très bien le faire de tout son cœur s'il y a un écart temporel, même minime, entre

24. *Ibid.*, p. 159.

25. L. Jaffro, « Cyrus' Strategy: Shaftesbury on Human Frailty and the Will », in Patrick Müller (dir.), '*New Ages, New Opinions': Shaftesbury in his World and Today*, Frankfurt, Peter Lang, 2014, p. 153-166.

26. Harry Frankfurt, *The Importance of What We Care About: Philosophical Essays*, Cambridge, Cambridge University Press, 1998, chap. 12.

le plaisir et sa désapprobation[27]. Certes, si cet écart temporel était très important, on passerait d'un cas de *weakness* à un cas de *recklessness*. Ceux qui comme Elster estiment douteux qu'il y ait une faiblesse de volonté synchronique estiment aussi que l'inversion de préférences dans la diachronie n'est pas toujours et est même rarement un cas de faiblesse de volonté. Mais cela ne remet pas en cause la thèse ici défendue.

À sa décharge, la conception de la liberté de la volonté que défend Harry Frankfurt est raisonnablement modérée, à la différence des conceptions extrêmes du libre-arbitre, sur deux points au moins :

(1) Elle ne fait pas du tout de la liberté de la volonté une condition nécessaire de la liberté de l'action et par conséquent de la responsabilité.

(2) Elle est compatible avec l'hypothèse d'une détermination causale ou tout aussi bien d'une détermination aléatoire des désirs de second ordre.

Cependant, la conception de Frankfurt comporte, il me semble, la même présupposition, incompatible avec la prémisse de la faiblesse constitutive, que les conceptions classiques – telle celle de Descartes – que Leibniz critiquait : elle suppose que des volitions de 2^{e} ordre peuvent avoir une influence directe et synchronique sur des volitions de 1er ordre.

Si l'on adopte la prémisse de la faiblesse constitutive, les techniques de contrôle de soi par anticipation, si elles semblent pouvoir convenir aux agents très faibles, dont l'irrationalité pratique est criante, voire pathologique (sans qu'ils perdent cependant l'identité personnelle), concernent également des agents aussi peu faibles que des agents humains en bonne santé peuvent l'être. Les uns comme les autres recherchent une forme d'intégrité pratique bien différente du *wholeheartedness* dont parle Frankfurt :

(1) En t_1 (moment initial de planification) les agents faibles comme Ulysse peuvent avoir les désirs qu'ils désirent avoir, mais alors que d'un point du vue frankfurtien cela constitue une condition suffisante pour que leur action soit *wholehearted*, de leur point de vue ce n'est pas le cas puisqu'ils sont préoccupés par le risque majeur qu'en t_2 (moment ultérieur de la décision en contexte difficile) cette intégrité soit perdue.

(2) En t_2, s'ils n'ont pas mis en place avec succès des techniques de contrôle par anticipation, ils ont d'autres désirs qu'ils désirent alors avoir, mais alors que d'un point du vue frankfurtien cela constitue encore une condition suffisante pour que leur action soit *wholehearted*, et qu'il en aille de même de leur point de vue alors, ce n'est pas le cas du point de vue qu'ils avaient en t_1.

(3) S'ils ont réussi à contrôler leurs désirs grâce à une ou plusieurs tactiques de contrôle par anticipation et avec un peu de chance, alors ils ont en t_2 les désirs qu'en t_1 ils désiraient avoir : leur action est-elle pour autant alors *wholehearted* ? Certainement pas, puisqu'ils ont préservé leur intégrité

27. Sur le caractère diachronique, dans certaines limites, de la faiblesse de volonté, voir Jeanette Kennett, *Agency and Responsibility. A Common-Sense Psychology*, Oxford, Oxford University Press, 2001, p. 54-61.

en luttant contre les désirs auxquelles ils se seraient identifiés en t_2 s'ils ne les avaient pas contrôlés indirectement.

La théorie de l'identification, de l'autonomie ou de l'intégrité que défend Frankfurt, paraît donc devoir être rejetée dans l'hypothèse de la faiblesse constitutive – qui est une sorte d'analogue psychologique de la doctrine de la Chute dans certaines théologies de la grâce (comme celles, aussi différentes soient-elles, de Blaise Pascal ou de Jean Calvin).

5. La fonction d'aide des engagements normatifs

Une question à l'horizon de cette étude : *certaines* de ces techniques de contrôle par anticipation mobilisent des engagements, dont la normativité peut adopter plusieurs formes et degrés, qu'il s'agisse des vœux, des promesses, des déclarations publiques (qui ne prennent pas nécessairement la forme de promesses et peuvent constituer une obligation de moindre force), des paris, ou bien sûr des contrats. *Certaines*, et pas toutes : détruire les cigarettes, le briquet, et les cendriers, est un exemple de technique de contrôle par anticipation, mais cette destruction est une contrainte et non pas un engagement normatif. Or la philosophie morale, politique et juridique – qui se présente bien souvent aujourd'hui sous l'intitulé « philosophie des normes » – comporte un biais en faveur du rôle spécifique des engagements normatifs qui tend à l'aveugler sur l'importance d'un autre type d'influence sur les conduites humaines, qui se combine généralement avec l'influence de l'engagement. Il faut remarquer ici que si certaines des techniques de contrôle par anticipation consistent non pas en des engagements, mais en l'organisation des circonstances qui sont susceptibles d'agir sur nos conduites (de manière causale – la notion de causalité qui est ici mobilisée n'est pas celle du physicien, mais plutôt celle du criminologue), on ne comprend pas pourquoi les autres techniques de contrôle indirect qui ont un aspect normatif agiraient purement à la manière dont le normatif influence les conduites humaines. La thèse selon laquelle les engagements normatifs, par exemple les promesses, ont un rôle *seulement* causal serait à l'évidence complètement absurde. Mais la thèse (le plus souvent implicite) selon laquelle ils n'ont aucun rôle causal me paraît être simpliste. Ils ont un rôle qui est *aussi* causal, à la manière dont l'institution de circonstances par un agent peut causer indirectement (ce n'est pas direct, puisque cela passe par les circonstances) un résultat qu'il ne parvient pas à atteindre directement. Lorsque je rappelle à quelqu'un la promesse qu'il m'a faite et qu'il n'a pas tenue, je ne suis pas en train seulement de le culpabiliser de ne pas avoir respecté sa parole, mais, dans certaines situations du moins, je l'incite à faire un effort supplémentaire par une pression qui instrumentalise cette promesse : même s'il ne tient pas spontanément parole, la promesse passée constitue un élément du contexte de ses choix présents, vers lequel je peux attirer son attention.

Reformulons cela en employant la distinction entre l'irréversibilité subjective et l'irréversibilité objective[28]. L'irréversibilité subjective est un engagement préalable par lequel un agent décide de se priver de certaines options de retour en arrière. L'irréversibilité objective est la disparition de ces options indépendamment des décisions de l'agent. Ma thèse est que certains engagements normatifs, et même les promesses, créent non seulement de l'irréversibilité subjective, mais une sorte d'irréversibilité objective dans la mesure où ils s'appuient sur la passivité de l'agent autant que sur sa volonté.

Il me semble que des agents faibles peuvent utiliser leurs engagements normatifs comme des « petits pas » (parfois plus efficaces que les « grands sauts » ou les ruptures) qui les entraînent là où ils veulent aller. À propos de ces petits pas, je ne suis pas entièrement convaincu par l'usage qui consiste à parler de *pre-commitment*, d'engagement préalable, pour désigner ce type d'appui, car cela pourrait suggérer que ces mesures sont *seulement* de l'ordre de l'engagement et s'harmonisent sans difficulté avec une conception forte de l'autonomie de la personne[29].

Par exemple, la promesse que je fais de ne pas accomplir telle action, en tant que telle, ne supprime pas cette action de l'ensemble de mes options accessibles ; c'est une irréversibilité subjective et non objective qu'elle introduit ; cependant la manière dont les agents considèrent les promesses ne s'épuise pas dans leur normativité propre ; la perspective du coût très élevé qu'il y aurait à trahir une promesse, tout comme simplement la pratique qui consiste à utiliser nos promesses comme une aide (l'aide est ici le côté positif de la contrainte) dans notre effort de maintien ou de transformation de l'identité pratique, suggèrent que l'option n'est pas fermée seulement en vertu de l'obligation que l'agent s'est donnée à lui-même auprès d'un tiers. Un engagement peut servir aussi à « brûler ses vaisseaux » ou à « couper les ponts ». Parfois, les deux aspects sont conjoints, mais distincts car portés par deux actions de l'agent concerné, l'une d'engagement, l'autre de contrainte ; par exemple, des pêcheurs signent un engagement de ne plus surpêcher telle espèce de poissons sur telle zone selon telle technique, et font un pas supplémentaire et d'une autre nature en changeant leurs équipements et techniques de pêche. Parfois, une seule et même action de la part de l'agent concerné porte la double fonction d'engagement et de contrainte. Ainsi une irréversibilité subjective initie une irréversibilité (finalement) objective. Un exemple de cela se trouve dans ce que j'appelle dans une autre étude la « tactique de l'île de Saint-Pierre », à propos de la résolution de Rousseau, relatée dans ses *Confessions* et dans les *Rêveries du promeneur solitaire*,

28. Sur laquelle Bénédicte Reynaud a attiré mon attention ; voir Patrick Cohendet, « Apprentissage organisationnel et cohérence : l'importance économique de la notion de réseau », in Bénédicte Reynaud (dir.), *Les limites de la rationalité*, tome II, *Les figures du collectif*, Paris, La Découverte, 1997, p. 71-98.

29. C'est manifeste sous la plume de Lubomira Radoilska, « Autonomy and Ulysses Arrangements », in L. Radoilska (dir.), *Autonomy and Mental Disorder*, Oxford, Oxford University Press, 2012, p. 252-280.

de se retirer sur le Lac de Bienne. Il s'agit d'une tactique qui institue volontairement les circonstances favorables à une transformation involontaire[30].

6. Conclusion

C'est aussi de cette manière, il me semble, que l'on peut donner une signification et une sorte de pertinence aux promesses faites à soi-même. Comme le remarque Schelling, les promesses à soi-même n'ont pas de force normative. Il l'explique par le fait que la personne à qui je promets, qui est habilitée à me libérer de ma promesse, dans ce cas, est moi-même[31]. Or une promesse que son auteur peut annuler à volonté n'est pas à proprement parler une promesse. Mais il faut faire observer que si dans le langage ordinaire il est bien question parfois de se faire une promesse à soi-même, ce n'est pas pour la force normative dont un tel engagement est dépourvu, mais pour l'aide qu'elle constitue. Je crois qu'il faut que la philosophie de la normativité prête davantage attention, dans un esprit plus humien que kantien, à la manière dont les agents s'appuient *aussi* sur leurs engagements normatifs, non comme sur les oracles d'une voix de la conscience, mais comme sur des béquilles.

Une éthique pour agents faibles accorde une place centrale à certaines caractéristiques métaphysiques qui restent généralement secondaires pour la philosophie morale. La temporalité de l'agir humain et son corrélat, qui est la diachronicité de la liberté pratique, comptent parmi ces caractéristiques qu'on aurait tort de négliger lorsqu'il s'agit de penser l'action morale.

L'ouverture interdisciplinaire de la philosophie à l'économie (et, à un bien moindre degré, l'ouverture réciproque de l'économie à la philosophie) que nous observons depuis quelques décennies est une opportunité pour le développement de cette éthique. En effet, l'économie peut nous aider à comprendre les problèmes de décision intertemporelle qui se posent à des agents dont les ensembles de préférences varient très fortement à travers le temps, ce qui met en péril leur effort de fidélité à certaines valeurs ou engagements. L'économie attire également notre attention sur l'importance des effets externes de l'action, y compris des effets externes qui peuvent être ceux d'une action d'un agent à l'égard d'une autre action ou d'un ensemble de préférences du même agent. C'est un aspect qui est très négligé par l'éthique des philosophes, alors qu'il concerne le vecteur de transformation morale le plus accessible.

30. L. Jaffro, « Comment produire le sentiment de l'existence ? », in Jean-François Perrin et Yves Citton (dir.), *Rousseau et l'exigence d'authenticité*, Paris, Classiques Garnier, 2014.

31. Schelling, *Choice and Consequence*, p. 99.

Il y a un prix à payer dans la conception d'une éthique pour agents faibles, qui a été signalé d'emblée : c'est nécessairement une éthique de l'hétéronomie, ce qui heurte les principes acquis dans une éducation kantienne ou aristotélicienne, et fait craindre une assimilation de l'éthique à une technique, à un art de vivre dont la valeur est instable parce que dépendante de la valeur des finalités qui sont poursuivies. Pour cette raison, une présentation moins incomplète de ce projet nécessiterait une exposition de la manière dont cette éthique essentiellement pratique s'articule avec une conception des finalités et des raisons morales[32]. Sans une telle conception, cette éthique n'est qu'une technique aveugle.

Pour donner un ou deux exemples des domaines dont l'approche pourrait être renouvelée par une éthique pour agents faibles, il suffit de penser au fait que ce que Michel Foucault appelait les techniques de l'existence ou les techniques de soi est généralement conçu comme un instrument d'une maîtrise de soi de niveau élevé. En réalité, comme on l'a vu avec l'analyse par Leibniz de la liberté diachronique de la volonté, ces techniques (et notamment celles qui ont recours à la retraite hors du monde ou bien au contrôle par un *coach*) constituent les seuls moyens de contrôle pour des agents qui n'ont pas une telle maîtrise de soi de niveau élevé.

Un autre domaine qui pourrait être abordé selon cette perspective est celui de ce qu'il est convenu d'appeler l'éthique appliquée. Cette appellation n'est pas satisfaisante parce qu'elle sous-entend que les problèmes d'une telle éthique sont des problèmes d'application pratique de conceptions déjà disponibles. En réalité, ce sont des problèmes d'action, très liés à l'état des technologies, qui eux-mêmes nous conduisent parfois à reconsidérer nos conceptions morales. Quoi qu'il en soit, il est absolument évident que de nombreux problèmes d'éthique appliquée qui nous préoccupent aujourd'hui, comme ceux de la fin de vie (par exemple la question de l'euthanasie), ou encore ceux de nos devoirs à l'égard des générations futures en matière de ressources naturelles et d'environnement, sont des problèmes de contrôle par anticipation. De ce point de vue, ce sont des problèmes qui se posent typiquement à des agents faibles ; et il est manifeste qu'ils sont faibles non pas simplement en raison de leur mauvaise volonté ou de leur négligence coupable, mais d'abord en vertu de la temporalité essentielle de la vie pratique[33].

32. Sur cette question, voir L. Jaffro, « Are Moral Reasons Response-Dependent? », *Philosophical Inquiries* (ETS), III, 2 (2015), p. 17-34.

33. Des versions de ce texte ont été présentées et discutées en janvier 2014 au séminaire « Les après-midis de Philosophie et Économie » (PHARE, Paris 1) ; en avril 2014 au forum organisé par l'Institut d'études avancées pour la culture européenne à l'Université Shanghai Jiao Tong ; et en octobre 2014 au 2e colloque international de philosophie économique, « Soi-même et autrui » (BETA, Strasbourg).

Bibliographie

Cohendet, Patrick, « Apprentissage organisationnel et cohérence : l'importance économique de la notion de réseau », in Bénédicte Reynaud (dir.), *Les Limites de la rationalité*, tome II, *Les figures du collectif*, Paris, La Découverte, 1997, p. 71-98.

Elster, Jon, *Le laboureur et ses enfants*, Paris, Minuit, 1987.

Feinberg, Joel, *Harm to Self*, New York, Oxford University Press, 1986.

Frankfurt, Harry, *The Importance of What We Care About: Philosophical Essays*, Cambridge, Cambridge University Press, 1998.

Jaffro, Laurent (dir.), *Croit-on comme on veut ? Histoire d'une controverse*, Paris, Vrin, 2013.

Jaffro, Laurent, « Cyrus' Strategy: Shaftesbury on Human Frailty and the Will », in Patrick Müller (dir.), '*New Ages, New Opinions': Shaftesbury in his World and Today*, Frankfurt, Peter Lang, 2014, p.153-166.

Jaffro, Laurent, « Comment produire le sentiment de l'existence ? », in Jean-François Perrin et Yves Citton (dir.), *Rousseau et l'exigence d'authenticité*, Paris, Classiques Garnier, 2014.

Jaffro, Laurent, « Are Moral Reasons Response-Dependent ? » *Philosophical Inquiries* (ETS), III, 2 (2015), p. 17-34.

Jaffro, Laurent, « Jekyll et Hyde. Quel modèle de multiplicité des soi ? », *Romanesques*, 7 (2015), p. 121-140.

Kennett, Jeanette, *Agency and Responsibility. A Common-Sense Psychology*, Oxford, Oxford University Press, 2001.

Korsgaard, Christine, *Self-Constitution. Agency, Identity, and Integrity*, Oxford, Oxford University Press, 2009.

Leibniz, *Nouveaux Essais sur l'entendement humain*, éd J. Brunschwig, Paris, Garnier-Flammarion, 1966.

Radoilska, Lubomira, « Autonomy and Ulysses Arrangements », in L. Radoilska (dir.), *Autonomy and Mental Disorder*, Oxford, Oxford University Press, 2012, p. 252-280.

Schelling, Thomas, « Self-Command in Practice, in Policy, and in a Theory of Rational Choice », *The American Economic Review*, 74, 2 (1984), p. 1-11.

Schelling, Thomas, *Choice and Consequence. Perspectives of an Errant Economist*, Cambridge, Mass., Harvard University Press, 1984.

Smith, Michael, *The Moral Problem*, Oxford, Blackwell, 1994.

Stocker, Michael, « Desiring the Bad: An Essay in Moral Psychology », *The Journal of Philosophy*, 76, 12 (1979), p. 738-753.

Williams, Bernard, « Formal Structures and Social Reality », in Diego Gambetta (dir.), *Trust. Making and Breaking Cooperative Relations*, Oxford, Blackwell, 1988, p. 3-13.