

UNE REVUE MULTIDISCIPLINAIRE SUR LES ENJEUX NORMATIFS
DES POLITIQUES PUBLIQUES ET DES PRATIQUES SOCIALES.

Les ateliers de l'éthique The Ethics Forum

A MULTIDISCIPLINARY JOURNAL ON THE NORMATIVE
CHALLENGES OF PUBLIC POLICIES AND SOCIAL PRACTICES.

- 4-27 Éthique et justice climatique : entre motivations morales et amORALES, **Pierre André** et **Michel Bourban**
- 28-64 La conservation autologue de sang de cordon ombilical : vers une nouvelle forme de participation biocitoyenne ?
Anouck Alary

DOSSIER : Metaethics, Normativity, and Value

- 65-69 Introduction, **Hichem Naar** et **Michele Palmira**
- 70-87 Goodness: Attributive And Predicative, **Michael-John Turp**
- 88-107 On-Conditionalism: On The Verge Of A New Metaethical Theory, **Toni Rønnow-Rasmussen**
- 108-132 Two Conceptions of Practical Reasons, **Christoph Hanisch**
- 133-150 A Stringent But Critical Actualist Subjectivism About Well-Being, **Stéphane Lemaire**
- 151-170 Les attitudes appropriées *verbatim*, **Fabrice teroni** et **Julien Deonna**
- 171-212 Le réalisme moral, **Peter Railton**

ÉTUDE CRITIQUE :

- 213-227 Réponse à une critique, **Francesco Paolo Adorno**

UNE REVUE MULTIDISCIPLINAIRE SUR LES ENJEUX NORMATIFS
DES POLITIQUES PUBLIQUES ET DES PRATIQUES SOCIALES.

Les ateliers de l'éthique The Ethics Forum

A MULTIDISCIPLINARY JOURNAL ON THE NORMATIVE CHALLENGES
OF PUBLIC POLICIES AND SOCIAL PRACTICES.

La revue est financée par le Conseil de recherches en sciences humaines (CRSH)
et administrée par le Centre de recherche en éthique (CRE)

The journal is funded by the Social Sciences and Humanities Research Council (SSHRC)
and administered by the Center for Research on Ethics (CRE)



Attribution 4.0 International

Vous êtes libres de reproduire, distribuer et communiquer les textes de cette revue au public, transformer et créer à partir du matériel pour toute utilisation y compris commerciale, selon les conditions suivantes :

- Vous devez créditer l'Œuvre, intégrer un lien vers la licence et indiquer si des modifications ont été effectuées à l'Œuvre. Vous devez indiquer ces informations par tous les moyens raisonnables, sans toutefois suggérer que l'Offrant vous soutient ou soutient la façon dont vous avez utilisé son Œuvre.

Pour tous les détails, veuillez vous référer à l'adresse suivante :
<https://creativecommons.org/licenses/by/4.0/legalcode>

You are free to copy and distribute all texts of this journal, transform, and build upon the material for any purpose, even commercially, under the following conditions:

- You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

For all details please refer to the following address:
<https://creativecommons.org/licenses/by/4.0/legalcode>

RÉDACTRICE EN CHEF/EDITOR

Christine Tappolet, Université de Montréal

COORDONNATEUR DE RÉDACTION/ADMINISTRATIVE EDITOR

Jean-Philippe Royer, Université de Montréal

COMITÉ EXÉCUTIF DE RÉDACTEURS/EXECUTIVE EDITORS

Éthique fondamentale : Natalie Stoljar, Université McGill

Éthique et politique : Ryoa Chung, Université de Montréal

Éthique et santé : Kristin Voigt, Université McGill

Éthique et économie : Peter Dietsch, Université de Montréal

Éthique et environnement : Gregory Mikkelsen, Université McGill

COMITÉ D'EXPERTS/BOARD OF REFEREES:

Arash Abizadeh, Université McGill

Charles Blattberg, Université de Montréal

Marc-Antoine Dilhac, Université de Montréal

Francis Dupuis-Déri, Université du Québec à Montréal

Geneviève Fuji Johnson, Université Simon Fraser

Axel Gosseries, Université catholique de Louvain

Joseph Heath, Université de Toronto

Julie Lavigne, Université du Québec à Montréal

Robert Leckey, Université McGill

Bruce Maxwell, UQTR

Pierre-Yves Néron, Université Catholique de Lille

Wayne Norman, Duke University

Hervé Pourtois, Université catholique de Louvain

Vardit Ravitsky, Université de Montréal

Mauro Rossi, Université du Québec à Montréal

Daniel M. Weinstock, Université McGill

NOTE AUX AUTEURS

Merci de soumettre votre article à l'adresse
jean-philippe.royer@umontreal.ca

Les articles seront évalués de manière anonyme par deux pairs. Les consignes aux auteurs se retrouvent sur le site de la revue (www.lecre.umontreal.ca/categorie/ateliers_ethique_ethics_forum/). Tout article ne s'y conformant pas sera automatiquement refusé.

GUIDELINES FOR AUTHORS

Please submit your paper to: jean-philippe.royer@umontreal.ca

Articles are anonymously peer-reviewed. Instructions to authors are available on the journal website (www.lecre.umontreal.ca/categorie/ateliers_ethique_ethics_forum/). Papers not following these will be automatically rejected.

ÉTHIQUE ET JUSTICE CLIMATIQUE : ENTRE MOTIVATIONS MORALES ET AMORALES

PIERRE ANDRÉ

DOCTORANT EN PHILOSOPHIE, UNIVERSITE PARIS-SORBONNE (PARIS IV)

MICHEL BOURBAN

COLLABORATEUR DE RECHERCHE, INSTITUT DE GEOGRAPHIE ET DURABILITE, UNIVERSITE DE LAUSANNE

RÉSUMÉ :

Dans un contexte d'urgence, les philosophes ne peuvent plus se contenter d'élaborer des théories idéales de la justice climatique fondées sur des motivations purement morales. Il est désormais nécessaire d'envisager des approches non idéales. Nous proposons ici de prendre au sérieux le problème de la motivation à l'action et nous mettons en avant certains motifs prudents pour lutter contre le changement climatique, en vue non pas de remplacer, mais de renforcer les motivations morales existantes, mais insuffisantes. Nous commençons par présenter trois grandes approches idéales qui ont prévalu jusqu'ici dans la recherche sur la justice climatique et sont fondées respectivement sur l'éthique déontologique, sur l'éthique des vertus et sur l'utilitarisme. Nous mettons en évidence leurs limites pratiques, en ciblant le problème de la motivation à l'action. Nous proposons ensuite quelques pistes pour résoudre ce problème en esquissant une approche non idéale de la justice climatique. Cette approche est centrée sur les motivations amoraux ressortant de l'examen attentif des perturbations systémiques globales et des contextes nationaux et locaux spécifiques. Elle est à notre sens incontournable pour motiver les États-nations, les entreprises et les individus. Il s'agit cependant moins de la substituer totalement aux théories idéales de la justice climatique que de compléter celles-ci pour établir une vision plus compréhensive qui distingue le problème de la justification morale de celui de la motivation.

ABSTRACT:

In a situation of urgency, philosophers can no longer rely exclusively on ideal theories of climate justice grounded on purely moral motives. It has become necessary to build non-ideal approaches. We propose to give serious consideration to the problem of motivation to act by putting forward some prudential reasons for tackling climate change—with the aim not of replacing existing but inadequate moral motivations, but of strengthening them. First, we identify three major ideal approaches that have so far prevailed in climate justice research, grounded respectively on deontological ethics, on virtue ethics, and on utilitarianism. We highlight their practical limitations, particularly with regards to the problem of motivation to act. Second, we open the way for an alternative nonideal approach to climate justice based on amoral motivations issuing from a careful examination of global systemic disruptions as well as of specific national and local situations. This approach cannot be ignored if one wishes to motivate nation-states, corporations, and individuals to act. However, our goal is not to replace ideal theory with nonideal theory, but rather to complete the former with the latter for a more comprehensive account of climate justice that distinguishes the problem of moral justification from the problem of motivation.

Le cinquième rapport du Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), paru entre fin 2013 et début 2014, est alarmant. Il confirme avec toujours plus de certitude que les êtres humains sont bien à l'origine d'une perturbation majeure et globale du climat en raison des émissions de gaz à effet de serre dans l'atmosphère. Il prévoit de graves conséquences de ces dérèglements climatiques sur la sécurité alimentaire, la santé, la mortalité et les migrations. Comme le quatrième chapitre du troisième groupe de travail le souligne, changement climatique et justice climatique sont désormais indissociables. Pourtant, que ce soit à l'échelle des négociations internationales entre États, des entreprises ou des individus, un changement radical de comportement à la hauteur de la gravité du problème se fait toujours attendre. L'Accord de Paris, malgré le progrès incontestable qu'il représente par rapport aux résultats des sommets mondiaux qui l'ont précédé, est, de l'aveu même des États qui l'ont adopté fin 2015, un accomplissement insuffisant pour parvenir à éviter une perturbation anthropique dangereuse du système climatique. Quelle position philosophique adopter face à cette situation d'urgence climatique?

Dans la recherche sur la justice climatique, les philosophes ont majoritairement laissé de côté la question de l'articulation entre la théorie et la pratique. Les positions les plus influentes tendent à se retrancher dans un débat entre des théories idéales qui peinent à penser le réel dans sa particularité, en construisant une approche normative déconnectée des données empiriques et des circonstances non idéales de la réalité. Si tous les philosophes ne tombent pas dans cet écueil,¹ la grande majorité des auteurs s'inscrit cependant dans une démarche qui a notamment le défaut de laisser soit entièrement, soit très largement de côté la question des motivations qui pousseraient les agents à respecter les principes de justice qu'elles ou ils défendent. Il est en ce sens nécessaire, pour élaborer une démarche qui cherche à penser la transition entre le réel tel qu'il est et le réel tel qu'il devrait être, de comprendre plus précisément en quoi les débats infinis sur les principes et les modèles théoriques qui leur servent de justification éloignent les philosophes des préoccupations concrètes des décideurs (citoyens et hommes et femmes politiques) et de l'urgence du problème climatique.

En se limitant à la justification des principes de la justice climatique, la plupart des positions se sont confrontées, sous des modalités différentes, au problème récurrent de la motivation à l'action, c'est-à-dire au problème de ce qui peut motiver les acteurs à lutter contre les injustices climatiques. Le problème de la motivation vient limiter la pertinence pratique de trois types de théories : les théories morales d'inspiration kantienne, qui supposent que le principal levier d'action est la conscience morale des individus, et les approches alternatives de l'éthique des vertus et de l'éthique utilitariste, qui prennent au sérieux le problème de la motivation, mais qui ne parviennent pas pour autant à atteindre le réel de façon efficace. Dans un premier temps, nous soulignons les limites propres à la démarche idéalisée des théoriciens de la justice climatique. Nous nous focalisons d'abord sur la position la plus influente dans le champ de recherche, celle de Simon Caney, fondée sur la conscience morale et les droits humains, pour montrer ensuite que les positions alternatives qui s'appuient sur la conscience environnementale des individus, comme celle de Dale Jamieson,

et sur l'altruisme, comme l'approche de Peter Singer, demeurent également trop idéales. Dans un second temps, nous proposons une approche philosophique qui peut être qualifiée de non idéale en ce qu'elle met en évidence les motivations amORALES qui complètent et renforcent les motifs moraux de la lutte contre le changement climatique. L'idée centrale est qu'il existe une convergence entre les prescriptions des principes de la justice et l'intérêt bien compris de celles et ceux à qui incombent les devoirs de justice climatique.

1. LES LIMITES DES THEORIES IDEALES DE LA JUSTICE CLIMATIQUE

La division du travail entre théorie idéale et théorie non idéale a été proposée par John Rawls et remonte au problème classique de la tension entre théorie et pratique tel qu'Emmanuel Kant l'avait formulé : ce qui vaut en théorie, compte tenu des exigences purement normatives de la raison, permet-il de penser ce qui se passe en pratique, compte tenu des circonstances imparfaites du monde? Si les agents idéaux sont systématiquement guidés par leur sens de la justice, qu'en est-il des agents réels, dont la raison morale est faillible? Cette division est aujourd'hui réinvestie par de plus en plus de philosophes, et il existe plusieurs manières de l'interpréter.² Ici, nous entendons par « théorie idéale » une approche analytique qui procède par idéalisation, modèles théoriques abstraits et expériences de pensée, et qui ne s'extrait du monde idéal qu'elle a construit qu'avec beaucoup de difficultés; par « théorie non idéale », nous entendons une approche pragmatique qui tient compte des circonstances non idéales du monde réel comme la faillibilité morale des individus. Les tenants d'une théorie idéale maintiennent les deux formes majeures d'idéalisation défendues par Rawls. Premièrement, l'obéissance aux principes de justice : les citoyens et hommes et femmes politiques sont des êtres raisonnables (capables d'un sens de la justice à l'aune de leur raison pratique) qui se conforment aux principes de justice. Deuxièmement, l'existence de circonstances favorables à la réalisation de la justice : les principes de la justice sont facilement applicables dans le monde réel.³ Les défenseurs d'une théorie non idéale montrent que cette double idéalisation se heurte à certaines caractéristiques imparfaites du monde réel, notamment aux ressources motivationnelles limitées des agents, dont la raison morale est limitée et faillible.

Une approche non idéale de la justice climatique comporte à notre sens deux tâches complémentaires. La première est de proposer des réformes institutionnelles à la fois justes et faisables, comme l'amélioration du fonctionnement des mécanismes de marché ou de taxation servant à réduire les émissions de gaz à effet de serre.⁴ La seconde tâche consiste à mettre en avant les motivations prudentielles qui renforcent les motivations morales que les actrices et acteurs économiques et politiques ont à déployer de telles réformes. Les deux tâches sont complémentaires au sens où la mise en œuvre de politiques climatiques justes et efficaces nécessite au préalable que les principaux acteurs politiques et économiques soient motivés à les adopter. Nous nous intéressons ici à la tâche qui se pose logiquement en premier : réduire le problème de la motivation à l'action contre le changement climatique. Pour reprendre l'idée mise en avant par

Ryoa Chung dans sa présentation du débat entre théories idéale et non idéale, il s'agit de développer une approche « pragmatique », tenant compte « de la convergence possible entre des finalités morales universalisables et les motivations prudentielles propres à chacun ».⁵

Pour le dire autrement, nous cherchons à montrer que l'éthique climatique devrait jouer un rôle plus important qu'elle ne le fait actuellement dans les débats de justice climatique. La littérature sur la justice climatique s'intéresse essentiellement aux principes de la justice, c'est-à-dire aux normes générales servant à guider les politiques de lutte contre le changement climatique et aux règles de distribution des coûts et bénéfices liés aux émissions de gaz à effet de serre. L'éthique climatique renvoie surtout quant à elle aux éthiques normatives auxquelles les philosophes du changement climatique ont recours pour justifier ces principes, comme le déontologisme, l'éthique des vertus et l'utilitarisme. Elle s'intéresse aussi aux problèmes de métaéthique comme la distinction entre considération morale et motivation effective à agir. Nous voudrions montrer qu'il existe une relation étroite et souvent négligée entre justice climatique et éthique climatique : la question du fondement de l'obéissance aux principes de justice climatique. L'obéissance stricte à ces principes suppose une motivation morale qui fait souvent défaut dans le monde réel, dans lequel c'est l'obéissance partielle, voire la désobéissance pure et simple, qui règne. Nous proposons une approche prenant au sérieux ce problème, moins pour contester que pour compléter les principaux résultats obtenus par les débats philosophiques sur le changement climatique. Il ne s'agit pas ici d'affirmer que l'ensemble des débats sur les principes de justice climatique distributive des théories idéales sont sans pertinence, mais au contraire de montrer que les deux approches ne sont pas irréconciliables, d'y voir deux points de vue normatifs différents, mais complémentaires sur le changement climatique, tout en soulignant les limites du premier type d'approche.

1.1. Les limites d'une approche déontologique : l'exemple de Simon Caney

Nombreux sont les principes de la justice climatique distributive qui sont justifiés, souvent implicitement, à partir d'une démarche kantiano-rawlsienne : pollueur-payeur,⁶ capacité à payer,⁷ bénéficiaire-payeur,⁸ sacrifice égal,⁹ non-coopérateur payeur¹⁰ ou encore non-gaspillage¹¹ sont autant de principes qui s'inscrivent dans le cadre d'une théorie idéale de la justice. Si, comme Rawls le précise dans sa *Théorie de la justice*, « [l]es principes de la justice sont analogues à des impératifs catégoriques au sens kantien »,¹² ces théories de la justice climatique reposent alors sur une éthique déontologique déduisant le devoir d'action d'un impératif catégorique. Autrement dit, ces positions philosophiques reposent entièrement sur la volonté bonne des agents à qui incombent les devoirs de justice climatique.

Le représentant le plus illustre de cette démarche est Simon Caney, qui défend une approche de la justice climatique centrée sur les droits humains,¹³ en s'inspirant des travaux de Thomas Pogge sur la justice globale.¹⁴ Comment cette posi-

tion est-elle construite? Caney soutient tout d'abord que les impacts du changement climatique menacent des droits humains fondamentaux. Il en isole trois en particulier pour appuyer son argumentation : les droits à la vie, à la santé et à la subsistance. Le droit à la vie est par exemple menacé par les événements climatiques extrêmes comme des ouragans plus violents et plus fréquents. Le droit à la santé est, quant à lui, confronté à la recrudescence de maladies vectorielles comme le paludisme. Enfin, le droit à la subsistance est mis en danger par une insécurité alimentaire accrue notamment par les sécheresses et les inondations. Il faut noter que Caney, dans le souci d'élaborer une position susceptible de recueillir un consensus le plus large possible, définit ces droits comme des droits négatifs (et non pas comme des droits-créances, plus exigeants), c'est-à-dire des droits à ne pas être privé, par une action humaine (comme celles à l'origine du changement climatique), de sa vie, de sa santé ou de ses moyens de subsistance. Sur la base de cette définition des droits, il déduit les devoirs qui y sont corrélés : devoirs de réduction des émissions de gaz à effet de serre, d'aide à l'adaptation aux effets du changement climatique qui adviendra malgré les réductions, et de compensation des violations des droits fondamentaux n'ayant pu être évitées en raison des mesures insuffisantes de réduction et d'adaptation.

Dans la tradition rawlsienne de la justice distributive, Caney s'intéresse ensuite à la question de la répartition équitable des devoirs de justice climatique en vue de prévenir et de compenser les injustices globales prenant la forme de violation des droits humains. C'est ici qu'intervient son raisonnement sur les principes. Caney s'intéresse en particulier à la justification morale d'un principe bien connu en droit international de l'environnement : le principe du pollueur-payeur. Il en fait ressortir les sources intuitives de légitimité (comme l'établissement d'une relation entre causalité et responsabilité morale), mais aussi les limites que constituent l'ignorance excusable, la non-responsabilité pour les émissions de ses aïeux et l'incapacité de certains pays en développement à assumer la responsabilité des émissions sans mettre en péril les droits de base de leurs habitants.¹⁵ Il propose, pour répondre à ces trois objections, un modèle hybride de justice distributive articulant le principe du pollueur-payeur au principe de la capacité à payer, qui attribue le fardeau de la lutte contre le changement climatique à celles et ceux qui ont la plus grande capacité à le supporter. Quand le principe du pollueur-payeur rencontre les limites mentionnées plus haut, c'est le principe de la capacité à payer qui prend le relai pour déterminer une distribution juste des devoirs de la justice climatique.

La position de Caney peut être qualifiée d'idéale pour deux raisons majeures. Premièrement, elle prolonge le questionnement sur les principes de la justice tout en conservant le type de démarche proposé par Rawls : elle demeure centrée sur la production de batteries de principes et, en ce sens, procède largement « *ex principiis* »¹⁶ L'objectif de ce théoricien de la justice climatique est d'orienter les politiques climatiques à partir de principes extérieurs aux situations non idéales. Au lieu de partir des données empiriques sur les inégalités, il se focalise sur la question des principes du juste, comme en témoigne sa prise de position dans le débat qu'il a engagé en posant la distinction entre l'isolationnisme et l'intégrationnisme. Tandis que la première position considère les principes de

justice climatique comme indépendants par rapport à une théorie générale de la justice, la seconde, à laquelle il se rattache,¹⁷ construit des principes normatifs à partir de modèles théoriques disponibles dans le champ de recherche de la justice globale (cosmopolitisme, suffisantisme, internationalisme pluraliste, etc.). La seconde limite caractérisant la position de Caney vient de ce qu'elle repose sur une conception partielle de la raison pratique laissant de côté les motivations prudentielles ou amORALES inhérentes à la rationalité instrumentale des actrices et acteurs. Le fait est que les motivations en jeu sont mixtes et même conflictuelles, ce qui explique souvent l'inaction. Il paraît donc nécessaire de développer une approche éthique pluraliste tenant compte de l'aspect limité et faillible de la raison morale en interrogeant la convergence possible entre des finalités morales universalisables et des motifs relevant de l'intérêt personnel. Dans ce qui suit, nous prenons en compte les réponses que d'autres théoriciens de la justice climatique ont apportées à la seconde limite; nous laissons de côté les réponses qu'il est possible d'apporter à la première et les approches philosophiques alternatives qu'il est possible de construire pour remplacer la démarche *ex principiis* par une démarche *ex datis*.¹⁸

1.2. Les limites de l'éthique des vertus : l'exemple de Dale Jamieson

Les théories de l'éthique climatique élaborées par Stephen Gardiner et Dale Jamieson¹⁹ nous aident à identifier les limites auxquelles se heurtent de manière récurrente les théories qui n'interpellent que la raison morale des agents. Tous les deux ont en effet reconnu l'importance du problème de la motivation en soulignant que ce sont les caractéristiques mêmes du changement climatique qui rendent exceptionnellement difficile la réalisation des devoirs de justice qu'il crée. Quatre raisons majeures poussent les agents à ignorer les prescriptions des principes de justice climatique distributive : la dispersion spatiale et temporelle des causes et des effets du changement climatique, qui en fait un phénomène global et largement différé; la répartition inégale de la vulnérabilité à ses impacts à travers la planète et entre les générations; l'inadéquation de nos institutions sociales, économiques et politiques à lutter contre un problème aussi profondément global et intergénérationnel; enfin les difficultés que rencontrent nos concepts normatifs classiques à détecter les injustices climatiques. Le changement climatique nous rend vulnérables à la « corruption morale », c'est-à-dire à la tentation d'échapper à nos devoirs de justice, et toute approche purement déontologique, comme celle de Caney, est pour cela insuffisante.²⁰

Pour Jamieson, le changement climatique concerne « la façon dont nous devrions vivre et dont les humains devraient se comporter avec leurs semblables et le reste de la nature ». ²¹ La nature du problème se trouve, selon lui, dans le système de valeurs de nos sociétés occidentales, notamment dans nos conceptions de l'intérêt personnel, de l'éthique et de la justice internationale. ²² Pour ne s'en tenir qu'à notre conception de la moralité, le changement climatique remet en question notre manière de concevoir le principe classique de non-nuisance : tandis que nous avons l'habitude de détecter des torts moraux dans les situations où un individu inflige une nuisance intentionnelle et identifiable à un autre individu, proche dans l'espace et le temps, dans le cas du changement climatique,

la nuisance est involontaire, difficile à identifier, diffuse dans l'espace et le temps, et d'origine collective. En conséquence, le changement climatique n'est pas perçu, en raison de nos concepts moraux conventionnels, comme un problème éthique grave et urgent. Quel type de solution faut-il alors mobiliser pour lutter contre ce nouveau type de problème moral?

Pour faire face à ces raisons qui poussent à l'inaction, Jamieson en appelle à une refonte du système de valeurs de nos sociétés occidentales. Selon lui, « [l]a seule manière de résoudre ce problème, qui correspond au problème d'action collective le plus large et le plus complexe au monde, est l'action d'un mouvement global de citoyens moralement motivés agissant comme un groupe d'intérêt politique extrêmement engagé ».²³ Ce mouvement serait composé de citoyens du monde qui célébreraient un mode de vie simple avec dignité et élégance, reposant sur des sources « naturelles » d'énergie, redécouvrant la nourriture, le plaisir de manger et les joies de vivre avec la nature. Pour parvenir à un tel résultat, Jamieson recommande de « développer une liste de vertus vertes » et d'« identifier les méthodes permettant de les inculquer de la manière la plus efficace possible ».²⁴ Cette liste contient par exemple l'humilité, la tempérance, l'amour, le souci de l'autre et le respect de la nature, et pourrait amorcer une révolution dans les mentalités.

Se retrouve ici le problème d'une position idéale : l'approche de Jamieson ne s'intéresse qu'aux motifs moraux de la lutte contre le changement climatique, sans prendre en compte la raison instrumentale des agents. Sa solution repose entièrement sur la conscience environnementale des individus. Or, dans notre monde non idéal, ce type de conscience est insuffisamment répandu : la plupart des gouvernements, des entreprises et des individus n'ont pas assez de sensibilité écologique pour que le projet d'un mouvement global de citoyens vertueux puisse être réalisé à temps et à l'échelle nécessaire. Et même chez ceux qui ont une telle conscience, certaines motivations amORALES, notamment celles qui relèvent de l'intérêt économique à court terme, sont susceptibles de les pousser à l'inaction. La convention constitutionnelle globale focalisée sur les générations futures récemment proposée par Gardiner²⁵ est soumise au même type d'objection : la préoccupation du sort des générations futures n'est sans doute pas suffisamment répandue pour que ce type de projet l'emporte sur la poursuite effrénée des bénéfices liés à l'exploitation des combustibles fossiles, à l'agriculture et à la déforestation. Même si ces auteurs parviennent à identifier le problème de la motivation, ils ne prennent pas non plus en compte l'aspect limité, faillible et imparfait de la raison morale lorsqu'ils proposent des solutions. Parier sur la conscience écologique des individus ne suffira pas non plus à motiver assez de citoyens, dirigeants d'entreprises et chefs d'État à lutter efficacement contre le changement climatique.

1.3. Les limites de l'utilitarisme : l'exemple de Peter Singer

Un des objectifs de l'éthique climatique d'inspiration utilitariste, représentée notamment par John Broome²⁶ et Peter Singer, est d'éviter les écueils des théories idéales de la justice climatique. Illustrons cette nouvelle position par une

analyse des travaux de ce dernier. Dans un premier texte,²⁷ Singer analyse plusieurs postures classiques des théories de la justice climatique, comme la responsabilité historique, le principe du pollueur-payeur et l'égalité stricte des droits d'émission par tête. Il montre en quoi elles peuvent être justifiées à partir de fondements utilitaristes, pour adopter finalement une approche égalitariste, qui est selon lui la plus à même d'augmenter la somme de bien-être global.²⁸ Dans un texte plus récent, Singer complète son analyse par une défense d'une éthique individuelle selon laquelle chacun est responsable des dommages causés par ses propres émissions.²⁹

Cette approche revêt ainsi plusieurs caractéristiques utilitaristes comme le conséquentialisme (que Singer s'efforce de concilier avec le fait que les comportements individuels pris séparément n'ont pas de conséquences perceptibles sur le climat) ou la règle « une personne, une voix », mise en exergue par la défense de l'égalitarisme strict. Mais ce sont avant tout les mentions du principe et du calcul d'utilité qui caractérisent sa position,³⁰ même s'il faut nuancer cette affiliation à une position strictement utilitariste dans la mesure où Singer souligne les difficultés extrêmes que rencontrerait la tentative d'un calcul de l'utilité globale dans le cas de la justice climatique. Face à des dynamiques physiques, économiques et politiques aussi complexes que celles qui sont à l'œuvre dans le cadre du changement climatique, la mise en pratique d'un calcul d'utilité global fiable se heurte en effet à des limitations importantes. Cette posture réaliste montre en quoi, dans son approche de l'éthique climatique, Singer cherche à se distancier des théories purement idéales. En effet, il s'efforce d'introduire dans son raisonnement les questions de faisabilité et d'efficacité qui font défaut aux deux premières positions étudiées plus haut. Singer montre ainsi les problèmes pratiques soulevés par l'égalité stricte par tête et la réduction colossale des émissions de gaz à effet de serre que cette norme impliquerait pour les pays industrialisés. Pour y répondre, il propose le recours à un marché global de droits d'émission doté de mécanismes de flexibilité qui simplifierait la transition économique des pays développés. Si mener à bien un calcul d'utilité semble irréaliste dans le cas du changement climatique, c'est néanmoins son approche conséquentialiste qui lui permet, contrairement au déontologisme et à l'éthique des vertus, de s'intéresser davantage au réel. Singer essaye ainsi de prendre en compte les conséquences supposées des actions qu'il propose, comme c'est le cas dans l'analyse économique qu'il fait des effets du mécanisme de droits échangeables.

Ces efforts pour se rapprocher d'une approche non idéale de la justice climatique ne paraissent toutefois pas suffisants sur trois points. Tout d'abord, la motivation à agir reste idéalement conçue puisque la maximisation du bien-être global ne semble pas constituer une source de motivation suffisante. Singer propose en effet l'altruisme comme levier fondamental de l'action, recommandant aux individus de se placer « du point de vue de l'univers » pour parvenir à réaliser « le plus grand bien possible »³¹ au niveau mondial. Ces motivations purement morales laissent de côté la question de l'intérêt des principaux individus, entreprises et États concernés par les politiques climatiques. Par ailleurs, même en ayant recours à l'outil que constituerait un marché global de droits

d'émissions, la défense d'un égalitarisme strict témoigne d'une deuxième forme d'idéalisation, car elle s'appuie sur des conditions favorables de réalisation de la justice : la mise en place d'un marché global de droits d'émissions pose en réalité de grands problèmes de faisabilité à court terme, comme en témoignent par exemple les difficultés du système européen de droits échangeables. Ce type de dispositif n'est pas totalement irréaliste, mais Singer ne s'interroge pas suffisamment sur les conditions de faisabilité d'un tel mécanisme. Enfin, la théorie utilitariste de la justice ne saurait, par nature, servir ni de base ni d'inspiration à une approche non idéale de l'éthique climatique, dans la mesure où elle constitue un système fondé sur un principe universel unique, l'utilité, qui s'applique sans distinction à toutes les organisations collectives et à tous les individus. Hautement spéculative, l'approche utilitariste reste une théorie idéale : l'objectif unique de maximisation de l'utilité abolit la possibilité d'une réflexion sur la justice climatique considérée pour elle-même, ce secteur n'ayant aucune autonomie par rapport au principe d'utilité.³²

2. VERS UNE APPROCHE NON IDEALE DE LA JUSTICE CLIMATIQUE

Ni l'approche déontologique de la justice climatique de Caney, ni celle de l'éthique climatique des vertus chez Jamieson, ni la position utilitariste de Singer ne parviennent véritablement à s'extraire des limitations d'une théorie idéale de la justice climatique. Le problème général de toutes ces positions est qu'elles reposent sur une conception métaéthique selon laquelle un jugement moral convaincant (comme la justification de l'existence d'injustices climatiques prenant la forme de violations des droits humains) implique forcément de lui-même une motivation à le respecter (comme par la mise en œuvre de politiques substantielles de réduction, d'adaptation et de réparation pour éviter ou réduire ces injustices). Autrement dit, la force motivationnelle des agents provient avant tout de facteurs internes aux considérations morales. Par contraste, on pourrait estimer que dans certains cas, les considérations purement morales, même si elles servent à qualifier les situations d'injustes, ne suffisent pas à motiver l'action; il faut également adjoindre des motivations extérieures aux considérations morales pour que les détentrices et détenteurs de devoirs respectent les exigences de la justice. De fait, il existe, dans le cas du changement climatique comme dans de nombreux autres problèmes moraux et politiques, un écart entre l'acceptation d'une règle et l'action qui s'y conforme. Les normes morales ne peuvent pas à elles seules garantir la conformité, de la part de l'agent moral, à l'action qu'elles prescrivent; tout ce qu'elles font, c'est recommander un certain type de comportement. Pour reprendre les termes de Dieter Birnbacher, « avoir des raisons morales pour agir et être effectivement motivé à agir sont deux éléments distincts, si bien qu'un mécanisme psychologique indépendant de l'acceptation de la règle morale est nécessaire pour que l'action soit conforme à cette dernière. »³³ La dissonance cognitive entre un comportement effectif et un comportement dont on est moralement partisan est un phénomène courant. On peut envisager de combler ce fossé de différentes manières, à l'aide d'institutions qui, par l'incitation ou la contrainte, augmentent la conformité aux règles morales, ou bien en mettant en avant d'autres motifs, extérieurs à la morale,

d'agir conformément à ce que prescrit la justice climatique. Si les deux approches tendent à se compléter, nous nous focalisons ici, comme annoncé ci-dessus, sur la seconde option.

Se référer exclusivement, pour combattre le changement climatique, à des motifs moraux comme les prescriptions d'un analogue de l'impératif catégorique, des vertus et de l'altruisme est largement insuffisant : d'autres facteurs relevant de l'intérêt personnel des individus doivent être intégrés pour parvenir à une approche non idéale de la justice climatique. C'est seulement ainsi qu'on peut esquisser une théorie non idéale de la justice climatique à même d'opérer la transition entre le réel tel qu'il est et le réel tel qu'il devrait être. Nous évoquons ici certains de ces motifs amoraux en nous focalisant d'une part sur les implications de l'existence de seuils critiques dans les systèmes artificiels et naturels et, d'autre part, sur un changement de perspective pour redescendre de l'échelle globale à l'étude des situations nationales, locales et individuelles concrètes. L'hypothèse faite ici est celle d'une convergence possible entre motivations prudentielles et motivations morales. Du point de vue global, la perspective de risques mondiaux n'épargnant personne et l'existence de points de basculement dans le système climatique sont des raisons fortes de lutter contre le changement climatique. Des points de vue nationaux et locaux, voire individuels, c'est toute une mosaïque de motifs prudents, dont nous esquissons ici un aperçu, qui peut entrer en résonance avec les motivations morales en faveur d'actions climatiques justes.

2.1. Les risques systémiques mondiaux

Un des sens du phénomène de la globalisation est que nous vivons dans un monde de plus en plus interconnecté et interdépendant. La vie sur notre planète est caractérisée par une inflation exponentielle des liens entre les personnes qui l'habitent. Nous vivons dans l'ère de l'« hyperconnectivité des systèmes »³⁴ techniques, sociaux et industriels, une période d'interaction spectaculaire des individus et des politiques. La vie quotidienne au XXI^e siècle se trouve entièrement branchée sur des macrosystèmes techniques : nos modes de vie reposent sur des technologies de plus en plus nombreuses et complexes qui assurent le fonctionnement de systèmes aussi variés que la défense militaire, la distribution de la nourriture et de l'eau et l'approvisionnement en électricité.³⁵ L'information, la finance, le tourisme, le commerce et ses chaînes d'approvisionnement, ainsi que les multiples infrastructures qui les sous-tendent, tous ces systèmes sont étroitement connectés. L'interdépendance est si forte que la perturbation d'un système peut causer la perturbation dans tous les autres systèmes qui y sont liés : des brèches dans nos infrastructures technologiques peuvent produire des effets dévastateurs. Aussi, les systèmes artificiels (dans le sens où ils sont le résultat de l'activité humaine) sont soumis à des effets de seuils : des perturbations en apparence minimales peuvent avoir des effets dévastateurs sur l'ensemble d'une région, d'un pays, voire du globe en entier. C'est ce que la crise du capitalisme financier du début de ce siècle illustre très bien. Le système financier international est devenu un réseau complexe de créances, d'obligations et de participations qui a

gagné en vitesse et en sophistication grâce aux transactions à haute fréquence, aux produits dérivés complexes, à la titrisation de la dette et à d'autres nouveaux outils financiers. Le prix à payer pour un tel système est que les chocs peuvent désormais se répandre très rapidement sur tout le réseau. C'est ce qui s'est passé en 2008, avec une crise bancaire qui ne s'est pas limitée au seul secteur financier : elle a affecté l'ensemble de l'activité économique en augmentant le chômage, en fragilisant la cohésion sociale et en sapant la confiance des consommatrices et consommateurs à travers le monde.

Mais il faut aller encore plus loin. Les systèmes financiers et la plupart des autres services reposent sur un socle physique : les réseaux d'infrastructures comme le transport routier, maritime, aérien ou ferroviaire, ainsi que les réseaux électriques et de télécommunications (dont Internet), qui sont eux aussi de plus en plus interconnectés et sophistiqués. Ces infrastructures, qui représentent « les grands piliers de nos sociétés »,³⁶ sont également soumises à des risques accrus dus à leur complexification. Par exemple, sans électricité, il est impossible de faire fonctionner les mines de charbon et les oléoducs, ainsi que de maintenir les systèmes de distribution d'eau courante, les chaînes de réfrigération, les systèmes de communication ou les centres informatiques et bancaires. De même, il suffit d'une infiltration par un pirate informatique dans le système d'un aéroport, d'une centrale nucléaire ou d'un système de défense d'une grande puissance militaire pour qu'une destruction à grande échelle devienne possible. Plus l'interdépendance des infrastructures est élevée, plus de petites perturbations peuvent avoir des conséquences importantes, voire catastrophiques. À l'âge des technologies de pointe, de petites ruptures de flux ou d'approvisionnement peuvent mettre en danger la stabilité du système global en provoquant des effets en cascade disproportionnés. La globalisation a transformé l'économie mondiale en un système hautement complexe qui augmente la vulnérabilité de chacun des systèmes.

Dans un monde marqué par un tel degré d'interconnexion, un nouveau type de risque émerge : le risque systémique mondial. C'est ce que le changement climatique illustre très bien. Par exemple, la sécurité sanitaire de tous est menacée par le spectre d'une épidémie globale. En déplaçant toujours plus de personnes dans des quartiers urbains insalubres et des camps de réfugiés sans infrastructures adaptées, comme un système d'égouts adéquat, un système de soins efficace et un accès à l'eau potable, le changement climatique augmente les risques d'une crise sanitaire mondiale en créant les conditions idéales pour l'apparition d'épidémies graves pouvant se propager à l'échelle mondiale. Si le changement climatique ne crée pas à lui seul ce risque, il augmente les chances qu'il se réalise en influençant les différents facteurs qui rendent une épidémie globale possible : il amasse des personnes dans des quartiers exigus et des camps de réfugiés avec très peu d'hygiène et de soins médicaux; il favorise le mouvement de personnes n'étant pas immunisées aux mêmes agents pathogènes; et il met les systèmes médicaux et les infrastructures sanitaires existants sous pression, en les poussant à saturation ou en les détruisant.³⁷ Le changement climatique n'est pas seulement extrêmement nuisible à la santé des pauvres du monde et des générations futures; il menace la santé de tout le monde, personne ne pouvant se mettre à l'abri d'une épidémie globale.

Dans un monde globalisé, nul n'est à l'abri des impacts du changement climatique. Certains des impacts du changement climatique pourraient aussi considérablement augmenter les risques de conflits violents à l'échelle nationale comme internationale dans certaines régions du globe. Si la globalisation est synonyme d'accroissement des entrelacements politiques, économiques, culturels et militaires, les formes de violences déclenchées par les impacts climatiques doivent être vues dans le contexte de cet entrelacement. D'un côté, le changement climatique exacerbe les risques posés par le terrorisme : la réduction de l'approvisionnement en eau, la baisse des rendements agricoles, l'augmentation du niveau des océans, ainsi que des déplacements de population toujours plus fréquents et importants contribuent à stimuler les tensions ethniques et religieuses existantes, à intensifier la compétition pour les ressources, à menacer la production et la distribution du pétrole, permettant ainsi au terrorisme de se répandre davantage.³⁸ D'un autre côté, le changement climatique contribue également à attiser le climat d'hostilité qui existe déjà dans certaines régions du monde et exacerbe les facteurs déclencheurs de conflits militaires dont les effets dévastateurs devraient largement dépasser les frontières des pays belligérants. Nous pouvons effectivement lire dans le dernier rapport d'évaluation du GIEC que « [l]e changement climatique peut accroître indirectement les risques de conflits violents – guerre civile, violences interethniques – en exacerbant les sources connues de conflits que sont la pauvreté et les chocs économiques (*degré de confiance moyen*). De multiples sources de données permettent de lier la variabilité du climat à ces formes de conflits ».³⁹

Par ailleurs, le changement climatique contribuera à déplacer des quantités massives de population qui chercheront à atteindre les îlots de prospérité et de stabilité que représentent pour eux l'Europe et l'Amérique du Nord. Cette perspective de gigantesques flux migratoires transnationaux représente un défi important pour les pays développés, qui ont déjà des difficultés à gérer la pression causée par les flux actuels de migrants sur leurs frontières. Selon Harald Welzer, le changement climatique « sera le plus grand défi de la modernité – en particulier face à l'inéluctable question du comportement à adopter vis-à-vis des masses de réfugiés, qui ne peuvent plus exister là d'où ils viennent et qui voudraient profiter des chances de survie offertes par les pays privilégiés ».⁴⁰ En 1995, lors de la Conférence de Berlin, Atiq Rahman, du Bangladesh Centre for Advanced Studies, a déclaré que « [s]i le changement climatique rend notre pays inhabitable, nous marcherons avec nos pieds mouillés dans vos salons ».⁴¹ Si cette mise en garde était déjà à prendre au sérieux en 1995, elle est encore plus pertinente vingt ans plus tard dans un monde où le niveau des océans augmente toujours plus rapidement à cause d'un réchauffement climatique toujours plus important.

2.2. Les points de basculement dans le système climatique

Si les risques systématiques globaux liés aux systèmes artificiels représentent de puissants motifs amoraux pour la lutte contre le changement climatique, ceux qui sont liés aux systèmes naturels sont des facteurs motivationnels au moins tout aussi importants. En plus des implications de la globalisation, nous souhaite-

rions mettre ici en évidence les implications de l'existence de seuils dans le système climatique, au-delà desquels certaines boucles de rétroaction capables de rendre le changement climatique autonome sont susceptibles de s'enclencher.

Selon la théorie scientifique des limites planétaires récemment développée par la Communauté des sciences du système Terre,⁴² la perturbation de certains systèmes biosphériques provoque des changements environnementaux capables de nous faire sortir de la stabilité de l'Holocène.⁴³ Cette théorie permet, en dressant une liste, de mettre au jour les trois systèmes planétaires qui ont été poussés au-delà de leurs limites critiques : le climat, le cycle de l'azote et la biodiversité. Si ces transgressions persistent (par exemple, si les concentrations atmosphériques de CO₂ ne sont pas rapidement stabilisées, puis réduites), la planète entière pourra être entraînée dans un nouvel état, dans lequel elle sera très probablement de moins en moins hospitalière avec les sociétés humaines.

La Terre est à concevoir comme un système complexe de limites fonctionnant comme un ensemble interdépendant : lorsqu'une limite est dépassée, cela peut avoir des impacts substantiels sur les autres systèmes. La perturbation du cycle climatique provoque des bouleversements dans d'autres systèmes, comme l'intégrité de la biosphère, ce qui en retour peut provoquer le dépassement d'autres limites, créant ainsi un cercle vicieux que personne ne maîtrise. Un élément décisif mis en lumière par la théorie des limites planétaires est que le comportement des systèmes biosphériques est sujet à des effets de seuil : au-delà d'un certain seuil de perturbation, leur état peut se modifier de manière abrupte et les mener vers un nouvel état, souvent avec des conséquences nuisibles, voire catastrophiques pour l'être humain et les autres êtres vivants. Par exemple, à partir d'une certaine concentration atmosphérique de gaz à effet de serre, le réchauffement climatique peut s'emballer, passant d'un modèle graduel à un modèle abrupt.

Quels sont ces seuils ou points de basculement? Tant la biosphère que les océans fonctionnent actuellement comme des puits de carbone. Au-delà d'un certain degré de réchauffement, ces éviers planétaires peuvent néanmoins devenir des sources substantielles de gaz à effet de serre. Par exemple, de larges quantités de dioxyde de carbone et de méthane se trouvant actuellement emprisonnées dans le pergélisol sibérien pourraient être émises dans l'atmosphère; de même, les océans pourraient émettre des quantités substantielles de méthane se trouvant actuellement congelées dans leurs profondeurs. Dans les deux cas, l'augmentation des températures globales se verrait radicalement amplifiée.⁴⁴ Dans son dernier rapport d'évaluation, le premier groupe de travail du GIEC confirme la possibilité d'un tel point de basculement, en précisant qu'il pourrait être atteint au fil du siècle si les émissions globales continuaient à augmenter.⁴⁵

Une notion centrale mobilisée par ce rapport est celle de rétroaction positive. Elle désigne une sorte de boucle de rétroaction, un enchaînement causal de facteurs, au sein d'un système écologique dynamique, dans lequel un facteur a un effet qui renforce en retour une de ses causes en amont, accélérant ainsi la transformation du système.⁴⁶ En l'occurrence, cette expression renvoie aux phénomènes causés par l'augmentation des températures, qui contribuent à leur

tour à amplifier le réchauffement, ce qui contribue ensuite à radicaliser les phénomènes en question, et ainsi de suite. Par exemple, lorsque les réserves de dioxyde de carbone et de méthane congelées se mettent à fondre en raison du réchauffement climatique, elles amplifient l'augmentation des températures, ce qui en retour amplifie le dégel de ces réserves de gaz à effet de serre, et ainsi de suite, donnant naissance à un cercle vicieux que personne ne maîtrise.

Le dégazage de méthane emprisonné dans le pergélisol sibérien et les fonds marins, la fonte accélérée des inlandsis du Groenland et de l'Antarctique Ouest, la fonte totale de la banquise arctique en été, le ralentissement de la circulation océanique thermohaline ou encore le dépérissement de la forêt amazonienne et des forêts boréales représentent autant de seuils dont le point de non-retour n'est pas encore connu, mais dont le franchissement pourrait faire basculer le climat vers un nouvel état inconnu jusqu'alors. Si certains de ces seuils peuvent s'enclencher d'ici quelques siècles, rien n'empêche que d'autres puissent s'enclencher dans le courant du XXI^e siècle. Selon le scientifique du système Terre Timothy Lenton et ses coauteurs, « une variété de points de basculement pourraient atteindre leur point critique au fil du siècle en raison du changement climatique anthropique ». Pour ces experts du climat, « [i]l est possible que nos sociétés se bercent dans un faux sentiment de sécurité provenant de projections rassurantes du changement global ».⁴⁷

Si l'humanité franchit ces points de non-retour, ce ne seront pas seulement les personnes pauvres éloignées et les membres des générations futures qui seront affectés; ce seront également les personnes riches ainsi que leurs enfants et petits-enfants, où qu'ils habitent. De nombreux intérêts des membres des classes moyennes et riches des pays développés et des nouveaux consommateurs des pays émergents comme la Chine, l'Inde, le Brésil et l'Afrique du Sud comptent ici dans la balance : leur bien-être économique, leur santé, leur subsistance et même leur vie peuvent être mis en danger dans un monde où les seuils critiques dans le système climatique seront dépassés. Lorsque les acteurs politiques (citoyens et hommes et femmes politiques) et économiques (producteurs et consommateurs) réfléchissent sur les politiques climatiques à mettre en œuvre dans les prochaines décennies, elles ont tout intérêt à prendre en compte ces nouvelles données des sciences du climat. Comme Ryoa Chung le souligne, dans la mesure où la diversité des motivations politiques et économiques qui animent les actrices et acteurs individuels et collectifs doit être reconnue, « il faut également accepter la nécessité de faire appel aux intérêts purement prudents des agents quand la raison morale ne suffit pas, en cas de motivations conflictuelles, à remporter la délibération ».⁴⁸

2.3. DÉPASSER LA JUSTICE GLOBALE

Ces risques globaux liés aux systèmes artificiels et naturels permettent de faire valoir des motivations fortes pour inciter les agents à lutter contre le changement climatique et ses effets. C'est ainsi qu'à l'échelle du système Terre, on trouve de nouveaux ressorts motivationnels différents du devoir, de la vertu et de l'altruisme, mais qui en même temps viennent renforcer leurs prescriptions

convergentes. On peut aussi découvrir de grandes sources de motivation en faveur de la justice climatique en redescendant, à l'inverse, de l'échelle globale à celles nationale, locale et individuelle. Une étude contextualisée des enjeux climatiques à l'échelle des États, des régions ou des individus met en évidence les contraintes du réel dans l'application des politiques climatiques. Une telle perspective de philosophie appliquée à des situations permet aussi de dévoiler des connexions pratiques entre le problème global et à long terme du changement climatique et des problèmes locaux et à court terme comme la pollution de l'air. Corrélativement, une analyse des situations politiques, sociales et économiques permet d'identifier des co-bénéfices locaux et à court terme des politiques de réduction des émissions, d'adaptation et de compensation. Ces autres bénéfices peuvent constituer des sources de motivation d'une autre nature que celles des motivations mises en avant par les théories idéales de la justice climatique. Cependant, ils ne peuvent être envisagés qu'à partir d'une étude minutieuse des cas réels aux échelles nationale et locale. Nous allons le montrer à travers quelques exemples qui illustrent des motivations susceptibles d'inciter les États-nations, les entreprises et les individus à lutter contre le changement climatique.

Un pays qui a sans conteste été au cœur des difficultés des négociations internationales sur le climat ainsi que des débats sur les principes de la justice climatique distributive est la Chine. Ce pays joue en effet un rôle central dans le changement climatique. Ses émissions ont plus que triplé entre 1990 et 2012, faisant de la République populaire de Chine le premier émetteur national de gaz à effet de serre à partir de 2006, devant les États-Unis.⁴⁹ Le négociateur en chef chinois pour les questions climatiques, Xie Zhenhua, a d'ailleurs admis en 2010 ce rôle de premier contributeur joué par son pays. Pour autant, le gouvernement chinois s'est opposé à de nombreuses reprises à tout accord faisant de la Chine le responsable principal de la lutte contre le changement climatique, arguant de la responsabilité historique des États-Unis pour les émissions produites depuis la Révolution industrielle, ainsi qu'en brandissant le principe des responsabilités communes mais différenciées, selon lequel on ne peut pas exiger d'un pays comme la Chine, qui doit encore lutter contre la pauvreté de centaines de millions d'habitants, le même engagement que celui qu'on attend des pays riches.⁵⁰ Les motivations morales à lutter contre le changement climatique afin de défendre, par exemple, les droits humains dans les pays les plus vulnérables se sont révélées sans grand secours pour pousser le gouvernement chinois à s'engager, dans la mesure où les États-Unis et les pays européens n'ont pas rempli au préalable leurs obligations. Certaines initiatives de justice ont certes été mises en place, comme le Fonds vert pour le climat, mais leurs résultats ne sont pas aujourd'hui encore suffisamment convaincants pour que l'on considère que les pays occidentaux s'acquittent de leur responsabilité et que la Chine doive désormais assumer les siennes. La portée des motivations liées à la justice reste donc limitée dans le cas de la Chine. Nous proposons alors de sortir de la perspective de la justice distributive globale pour descendre à l'échelle nationale afin d'étudier le contexte réel de la Chine et de montrer qu'il existe d'autres motivations pour ce pays de promouvoir indirectement la justice climatique.

Ce que l'on observe alors en particulier, ce sont les niveaux très préoccupants de pollution de l'air sur le plan local, notamment dans les grandes villes. Il s'agit là d'un phénomène différent du changement climatique. La pollution est locale et non différée dans le temps, même si elle a des origines communes avec les dérèglements climatiques : la généralisation de l'automobile et le recours aux centrales à charbon entre autres. Or, la pollution de l'air fait aujourd'hui grand débat en Chine. L'agitation provoquée par le documentaire de Chai Jing, « La Chine dans la brume : sous le dôme » en témoigne. Ce reportage dénonçant les niveaux extrêmes de pollution de l'air et leurs graves conséquences sur la santé a été regardé plus de 200 millions de fois sur Internet en Chine avant d'être censuré par les autorités.⁵¹ Aujourd'hui, la réduction de la pollution, notamment par la fermeture des centrales à charbon, constitue un débat de grande ampleur, remettant en cause la croissance économique des dernières décennies. Il est donc dans l'intérêt du gouvernement chinois, soucieux de la stabilité de son régime politique, de lutter contre les causes de la pollution de l'air, sans parler des coûts économiques causés par le secteur de la santé. Ainsi trouve-t-on dans l'examen du contexte chinois des motivations prudentielles qui peuvent inciter l'État chinois à lutter à la fois contre les causes d'une pollution *hic et nunc*, et contre la source du changement climatique. Les motivations amORAles relevant de l'intérêt national viennent renforcer les motivations morales à lutter contre le problème climatique.

Plus généralement, la combustion des énergies fossiles n'a pas seulement des impacts négatifs globaux à long terme; elle a également des effets néfastes locaux à court terme. Certains gaz à effet de serre comme le carbone noir, le méthane et certains chlorofluorocarbures (CFC) qui ont une durée de résidence dans l'atmosphère nettement inférieure à celle du dioxyde de carbone sont émis par des secteurs très précis de l'économie, et les coûts liés à leur réduction sont nettement moins élevés que ceux liés aux émissions de CO₂. Qui plus est, la réduction de ces émissions fournit des bénéfices locaux en matière de santé, d'agriculture et de développement humain. L'action contre les émissions de méthane et de carbone noir pourrait permettre d'éviter 4,7 millions de décès prématurés d'ici 2030 et d'améliorer sensiblement les rendements agricoles. Ces bénéfices seraient avant tout locaux : par exemple, la réduction des émissions de méthane et de carbone noir provenant de la région Arctique et des pays voisins pourrait ainsi permettre d'éviter 47 800 décès prématurés dans la même région. Ces opportunités de réduction à faibles coûts et à hauts bénéfices permettraient en outre de réduire d'environ 0,5°C l'augmentation du réchauffement global entre 2010 et 2040, certains gaz comme le méthane ayant un forçage radiatif nettement plus élevé que le dioxyde de carbone.⁵²

Cherchant à changer de perspective pour dépasser l'approche de la justice globale en s'intéressant aux contextes particuliers, on trouve d'autres exemples de motivations prudentielles à l'action contre le changement climatique. C'est par exemple le cas des motivations économiques à développer des secteurs d'activités contribuant à construire une économie décarbonée comme celui des énergies renouvelables. Pour les investisseurs et les industriels, le marché des énergies renouvelables représente en effet désormais une option rentable. Pour

preuve, les investissements dans les énergies renouvelables ont mondialement atteint 318 milliards de dollars en 2014, soit une croissance de 19% par rapport à 2013.⁵³ Dans de nombreuses régions du monde, les coûts des énergies éolienne et solaire concurrencent désormais ceux des sources classiques d'énergie. Comme l'économiste Nicholas Stern le souligne dans une contribution récente, « le coût de l'énergie pouvant être fournie par ces dispositifs est compétitif (c'est-à-dire sans qu'il n'y ait besoin de subsides ou de taxes carbone) dans environ 80 pays ».⁵⁴ Ces opportunités de marché toujours plus nombreuses représentent des motivations amORAles fortes pour les agents économiques. Elles sont susceptibles de participer à la transition vers une économie décarbonée et donc de lutter contre le changement climatique.

Ces deux éléments – les bénéfices locaux à court terme liés à la réduction des émissions des gaz à effet de serre et la chute du prix des énergies renouvelables – montrent que la transition énergétique nécessaire pour éviter une perturbation anthropique dangereuse du système climatique n'est pas aussi coûteuse que ce que les lobbies de l'industrie des combustibles fossiles voudraient nous le faire croire. L'effet combiné de ces deux facteurs illustre en fait des bénéfices nets à court et moyen termes associés à la lutte contre le changement climatique. Autrement dit, les coûts absolus de la transition de l'économie mondiale à des sources d'énergie renouvelable sont en train de décliner, tandis que les bénéfices directs associés à la réduction des émissions de gaz à effet de serre deviennent plus manifestes. La structure collective du problème climatique, selon laquelle il est collectivement rationnel de coopérer pour réduire les émissions mondiales, mais individuellement irrationnel de le faire, est pour ces raisons partiellement en train de s'effriter. Cela nous donne des raisons d'espérer qu'un changement climatique dangereux peut encore être évité, mais seulement à condition que les acteurs économiques et politiques locaux soient conscients des co-bénéfices liés à la lutte contre le changement climatique.⁵⁵ De ce point de vue, le mouvement social pour la justice climatique, notamment les campagnes de la société civile comme le mouvement pour le désinvestissement des énergies fossiles, joue un rôle central.⁵⁶

On peut enfin mettre en évidence des motivations prudentielles chez les individus, qui concourent indirectement à lutter contre le changement climatique. Le cas de la consommation de viande rouge présente un exemple de ce genre de motivations. L'Organisation mondiale de la santé (OMS) a récemment annoncé que la viande rouge était cancérigène, ce qui peut constituer une forte motivation pour les individus des pays industrialisés à réduire leur consommation de viande bovine.⁵⁷ Or, l'élevage industriel ou intensif, au premier rang duquel l'élevage bovin, est à l'origine d'une part significative des émissions mondiales de gaz à effet de serre (environ 18%, c'est-à-dire plus que les émissions provenant du secteur des transports) en raison du méthane produit, un gaz au pouvoir de réchauffement vingt fois supérieur à celui du dioxyde de carbone.⁵⁸ Un changement d'habitudes alimentaires motivé par des raisons égoïstes liées à la santé individuelle participe de la lutte contre le changement climatique. C'est ainsi une attitude prudentielle envers les maladies chroniques, sur le plan individuel, qui entre en jeu. Ce type de motivation est particulièrement adapté pour inciter

les populations riches des pays développés et émergents, soucieuses de leur bonne santé, à changer leur comportement.

De ce point de vue, le végétarisme représente un engagement individuel triplement bénéfique. Premièrement, il permet de réduire la souffrance animale en réduisant les traitements abjects et immoraux infligés aux animaux dans le cadre de l'élevage industriel. Les deux préoccupations essentielles de ce type d'élevage très répandu sont simples : maximiser la production et minimiser les coûts. Les effets abominables des traitements qui découlent de cette logique sont suffisamment documentés pour qu'il ne soit pas nécessaire de s'y arrêter ici.⁵⁹ Deuxièmement, le végétarisme promeut également la santé humaine. L'élevage industriel nuit à la fois aux animaux et aux humains, notamment du point de vue de la sécurité alimentaire : les produits que l'on donne aux animaux comme les hormones de croissance et les antibiotiques se retrouvent dans la viande que l'on mange, et certaines de ces substances sont cancérigènes. De plus, l'élevage industriel à l'ère de la mondialisation implique des épizooties et augmente le risque de pandémie. Ici aussi, le risque sanitaire concerne tout le monde : la grippe porcine, la « vache folle », la fièvre aphteuse et la grippe aviaire représentent autant d'effets pervers du traitement infligé aux animaux de consommation.⁶⁰ Enfin, le végétarisme contribue également à la réduction des coûts environnementaux. D'un côté, produire de la viande pollue les sols, l'air et l'eau, participant entre autres à l'érosion de la biodiversité. D'un autre côté, l'élevage des ruminants, l'utilisation d'engrais de synthèse et la déforestation servant à convertir les forêts en pâturages causent des émissions massives de gaz à effet de serre dans l'atmosphère. La déforestation cause principalement des rejets de CO₂, les pratiques agricoles des émissions de méthane, et l'utilisation d'engrais des émissions de protoxyde d'azote.

Le cas du végétarisme est un exemple éloquent du pluralisme motivationnel auquel on peut avoir recours pour promouvoir la lutte contre le changement climatique : on y observe une convergence entre des motifs associés à nos devoirs envers les plus pauvres du monde et les générations futures, à l'éthique environnementale, à l'éthique animale et à notre intérêt égoïste à vivre en bonne santé. L'étude de ce genre de convergences doit être davantage approfondie afin de montrer plus largement en quoi l'examen attentif des contextes nationaux, locaux et individuels permet ainsi d'identifier, à travers des connexions avec d'autres problèmes non différés dans le temps et dans l'espace, des motivations nouvelles en faveur de la lutte contre le changement climatique.

3. CONCLUSION : IMPLICATIONS POUR UNE APPROCHE NON IDÉALE DE LA JUSTICE CLIMATIQUE

L'approche proposée ici permet de se faire une idée plus complète de la diversité des intérêts mis en danger par le changement climatique, et surtout de réaliser qu'il peut aller dans le sens de l'intérêt des États, des entreprises et des individus de respecter leurs devoirs de justice climatique. Une perturbation anthropique dangereuse du système climatique ne bénéficierait à personne; au contraire, tout le monde en souffrirait, d'une manière ou d'une autre. De plus,

une analyse des situations nationales, locales et individuelles permet de mettre en évidence les co-bénéfices des politiques climatiques, bénéfiques d'une nature différente, non différés dans le temps et l'espace.

L'approche de la motivation qui a prévalu jusqu'ici dans les théories de la justice climatique, une approche idéale fondée sur l'impératif catégorique, la vertu ou l'altruisme, n'a pas examiné ces motivations amORALES pourtant fortes. Or, ces dernières doivent impérativement être prises en compte pour opérer un changement de comportement à grande échelle, de toute urgence. Nous proposons donc, en prenant le parti de distinguer le problème du fondement du jugement éthique et celui de la motivation à l'action, une approche non idéale qui s'appuie sur des motifs prudentiels. Les théoriciens de la justice climatique ont raison d'insister sur les multiples motifs moraux pour lutter contre le changement climatique; cependant, ils ont jusqu'ici négligé tous les motifs amORAUX qu'il y a à le faire, et nous souhaitons remédier à cette omission.

Nous sommes tout à fait conscients que la distinction entre théorie idéale et théorie non idéale ne se réduit pas au problème de la motivation, mais inclut aussi la question des institutions politiques et sociales à même de pousser les agents moraux à se conformer à leurs devoirs, notamment par la contrainte et par des incitations. Il est vrai que la question institutionnelle constitue elle aussi une piste privilégiée pour répondre au problème de l'inaction et de l'inertie politique. Ce problème étant de taille, il ne pouvait cependant être traité ici en profondeur en complément de celui de la motivation à l'action. Nous nous sommes intéressés ici au problème de la motivation, car il nous semble être prioritaire par rapport à celui des institutions, du moins dans le cadre d'une prise de décision démocratique et pour une raison simple : sans motivation des citoyens et des hommes et femmes politiques à réformer les institutions, il n'y a pas de réforme possible. Si la question de la réforme des institutions est d'une grande importance (notamment dans la prise en compte du long terme qui leur fait tragiquement défaut), elle nous semble n'intervenir que dans un second temps. De ce point de vue, notre approche non idéale de la justice climatique mérite encore d'être complétée. Une autre limite possible à l'approche défendue ici est que la question des motivations prudentielles à agir a déjà été étudiée dans d'autres secteurs de la justice, sans pour autant apporter de résultats probants quant à la sortie de l'inertie politique. Pourquoi en serait-il autrement dans le cadre de la justice climatique? Nous avons soutenu l'hypothèse selon laquelle le problème majeur auquel était confrontée la justice climatique tenait à la question de la motivation, ce qui n'est pas nécessairement le cas d'autres domaines de la justice. Dans la mesure où certains exemples étudiés plus haut comme le végétarisme, la lutte contre les pollutions locales, ou encore l'essor économique des énergies renouvelables témoignent d'une forme de convergence attestée entre des motivations prudentielles et les devoirs de justice climatique, cette voie peut contribuer à faciliter la transition entre la théorie et la pratique – du moins il n'est pas interdit de l'espérer.

Une dernière limite de l'approche proposée ici tient aux conséquences potentiellement négatives du développement d'une éthique climatique reposant sur des motivations prudentielles. On pourrait en effet lui reprocher de faire la promotion d'une forme de moins-disant moral par rapport aux autres théories étudiées plus haut. Un tenant de l'éthique des vertus rétorquerait ainsi peut-être que fonder la motivation sur les intérêts, parfois égoïstes, des agents, revient à promouvoir le développement de certains vices alors que la lutte contre le changement climatique était justement l'occasion de faire renaître d'antiques vertus oubliées aujourd'hui. Des partisans de l'éthique déontologique déploreraient certainement, quant à eux, l'occasion manquée de faire advenir une véritable conscience cosmopolitique globale du devoir, rendue plus que jamais urgente et, en même temps, à notre portée. Ces arguments ne manquent pas de force et nous ne pouvons que nous associer à ce type de projets en général, visant un progrès moral de l'humanité. Néanmoins, l'examen de l'inertie politique actuelle, des risques d'emballement du système climatique et de l'échec, jusqu'à présent, des exigences purement morales à réformer nos comportements nous amène à placer en priorité par rapport à ce type d'exigences la survie d'une humanité moins inégalitaire qu'elle ne le serait si le changement climatique s'accroissait.

Nous voudrions insister pour terminer sur le fait qu'il ne s'agit pas de se débarrasser des théories idéales de la justice climatique, mais plutôt de les compléter dans une perspective de philosophie appliquée aux situations réelles et à l'urgence. Plutôt que d'adopter une approche purement prudentielle, comme celle récemment défendue par John Broome,⁶¹ voire une approche réaliste fondée sur une version trop simpliste des relations internationales, comme celle d'Eric Posner et de David Weisbach,⁶² notre objectif est plutôt d'adopter une éthique pluraliste, consciente des limites des approches exclusivement morales, mais qui maintient en même temps les acquis des vingt années de débats en justice et en éthique climatiques. Il s'agit de montrer non seulement que les différentes éthiques climatiques sont largement conciliables, mais également qu'une approche prudentielle obtient des résultats similaires venant renforcer les motifs moraux. Les approches déontologique, utilitariste, des vertus et de l'intérêt bien compris convergent, dans le cas du changement climatique, vers le même résultat dans la pratique : il faut mettre en œuvre des politiques climatiques drastiques non seulement parce qu'il s'agit d'un devoir moral, parce que cela correspond à un comportement vertueux, et parce que cela maximise l'utilité à l'échelle globale, mais aussi parce qu'il est dans nos intérêts de le faire.

REMERCIEMENTS

Pour leurs commentaires très utiles qui nous ont aidés à améliorer sensiblement cet article, nous remercions les évaluatrices et évaluateurs anonymes de la revue et Lisa Broussois.

NOTES

- ¹ Pour une exception, voir Eckersley, Robyn, « Moving Forward in the Climate Negotiations: Multilateralism or Minilateralism? », *Global Environmental Politics*, vol. 12, n° 2, 2012, p. 24-42.
- ² Voir notamment Simmons, A. John, « Ideal and Non-ideal Theory », *Philosophy & Public Affairs*, vol. 38, n° 1, 2010, p. 5-36; Valentini, Laura, « Ideal vs. Non-ideal Theory : A Conceptual Map », *Philosophy Compass*, vol. 7, n° 9, 2012, p. 654-664; Chung, Ryoa, « Théories idéale et non idéale », in Jeangène Vilmer, Jean-Baptiste et Ryoa Chung, (dir.), *Éthique des relations internationales. Problématiques contemporaines*, Paris, PUF, 2013, p. 63-91.
- ³ Chung, Ryoa, « Théories idéale et non idéale », in Jeangène Vilmer et Chung, (dir.), *Éthique des relations internationales. Problématiques contemporaines*, p. 65.
- ⁴ Voir sur ce point : Bourban, Michel, « Vers une éthique climatique plus efficace : motivations et incitations », *Les ateliers de l'éthique/The Ethics Forum*, vol. 9, n° 2, 2014, p. 4-28.
- ⁵ Chung, Ryoa, « Théories idéale et non idéale », in Jeangène Vilmer et Chung, (dir.), *Éthique des relations internationales. Problématiques contemporaines*, p. 82. Pour une présentation de la distinction entre théorie idéale et non idéale également centrée sur le problème de la motivation à l'action, voir Bourban, Michel et Simone Zurbuchen, « Justice globale », in Hervouët, François, Mbongo, Pascal et Carlo Santulli, (dir.), *Dictionnaire encyclopédique de l'État*, Paris, Berger-Levrault, 2014, p. 579-584, p. 583.
- ⁶ Shue, Henry, « Global Environment and International Inequality » (1999), in Gardiner, Stephen M., Caney, Simon, Jamieson, Dale et Henry Shue, (dir.), *Climate Ethics: Essential Readings*, Oxford, Oxford University Press, 2010, p. 101-111.
- ⁷ Caney, Simon, « Climate Change and the Duties of the Advantaged », *Critical Review of International and Social Philosophy*, vol. 13, n° 1, 2010, p. 203-228.
- ⁸ Page, Edward A., « Climatic Justice and the Fair Distribution of Atmospheric Burdens: A Conjunctive Account », *The Monist*, vol. 94, n° 3, 2011, p. 412-432.
- ⁹ Miller, David, « Global Justice and Climate Change: How Should Responsibilities Be Distributed? », in Peterson, Grethe B., (dir.), *The Tanner Lectures on Human Values*, Salt Lake City, University of Utah Press, 2009, p. 119-156.
- ¹⁰ Symons, Jonathan, « The Non-cooperator Pays Principle and the Climate Standoff », in Harris, Paul, (dir.), *China's Responsibility for Climate Change: Ethics, Fairness and Environmental Policy*, Bristol, Policy, 2011, p. 99-120.
- ¹¹ Cosson-Eide, Hans, « The Wastefulness Principle. A Burden-sharing Principle for Climate Change », *Journal of Global Ethics*, vol. 10, n° 3, 2014, p. 351-368.
- ¹² Rawls, John, *Théorie de la justice* (1971), trad. C. Audard, Paris, Seuil, 2009, p. 289.
- ¹³ Caney, Simon, « Climate Change, Human Rights, and Moral Thresholds », in Gardiner, et coll., (dir.), *Climate Ethics*, p. 163-177.
- ¹⁴ Simon Caney écrit dans l'article que nous venons de citer : « [m]on approche ici est redevable à celle développée par Thomas Pogge dans ses travaux pionniers sur la pauvreté globale. » (*Ibid.*, p. 175, n23).
- ¹⁵ Caney, Simon, « Climate Change and the Duties of the Advantaged ».
- ¹⁶ Nous reprenons cette expression à Alain Renaut et Jean-Cassien Billier, qui la reprennent eux-mêmes à Kant : Renaut, Alain et Jean-Cassien Billier, « Pour une approche philosophique globale des inégalités », *Socio*, vol. 4, mars 2015, p. 241-264, p. 245.

- ¹⁷ Caney, Simon, « Just Emissions », *Philosophy & Public Affairs*, vol. 40, n° 4, 2012, p. 255-300; Caney, Simon, « Distributive Justice and Climate Change », in Olsaretti, Serena, (dir.), *The Oxford Handbook of Distributive Justice*, Oxford, Oxford University Press, à paraître.
- ¹⁸ Nous développons une telle démarche dans les deux études applicatives suivantes : André, Pierre et Geoffroy Lauvau, « Entre responsabilité et vulnérabilité extrêmes : la question du changement climatique en Chine », in Renaut, Alain et coll., *Inégalités entre globalisation et particularisation*, Paris, PUPS, 2016, p. 613-629; Bourban, Michel et Geoffroy Lauvau, « Les petits États insulaires : peut-on compenser la perte d'un territoire? », in Renaut, Alain et coll., *Inégalités entre globalisation et particularisation*, p. 631-649.
- ¹⁹ Gardiner, Stephen M., *A Perfect Moral Storm: The Ethical Tragedy of Climate Change*, Oxford, Oxford University Press, 2011; Jamieson, Dale, *Reason in a Dark Time: Why the Struggle Against Climate Change Failed – and What it Means for our Future*, Oxford, Oxford University Press, 2014.
- ²⁰ Gardiner, Stephen M., « Human Rights in a Hostile Climate », in Holder, Cindy et David Reidy, (dir.), *Human Rights : The Hard Questions*, Cambridge, Cambridge University Press, 2013, p. 211-230; Jamieson, *Reason in a Dark Time*, p. 152-160.
- ²¹ Jamieson, Dale, « Ethics, Public Policy, and Global Warming » (1992), in Gardiner, Stephen M. et coll. (dir.), *Climate Ethics*, p. 77-86, p. 79.
- ²² Jamieson, Dale, « Energy, Ethics, and the Transformation of Nature », in Arnold, Denis G. (dir.), *The Ethics of Global Climate Change*, Cambridge, Cambridge University Press, 2011, p. 16-37, p. 30.
- ²³ *Ibid.*, p. 36.
- ²⁴ Jamieson, Dale, « When Utilitarians Should Be Virtue Theorists » (2007), in Gardiner, Stephen M. et coll., (dir.), *Climate Ethics: Essential Readings*, p. 315-331, p. 326. Voir également : Jamieson, Dale, « Respecter la nature », trad. J. Chalier, *Esprit*, vol. 12, déc. 2015, p. 23-33.
- ²⁵ Gardiner, Stephen M., « A Call for a Global Constitutional Convention Focused on Future Generations », *Ethics & International Affairs*, vol. 28, n° 3, 2014, p. 299-315.
- ²⁶ Broome, John, *Climate Matters: Ethics in a Warming World*, New York et Londres, W. W. Norton & Company, 2012.
- ²⁷ Singer, Peter, *One World. The Ethics of Globalization*, 2^e éd., New Haven et Londres, Yale University Press, 2004, chap. 2 : « One atmosphere », p. 14-50.
- ²⁸ *Ibid.*, p. 43.
- ²⁹ Singer, Peter, *Practical Ethics*, 3^e éd. révisée, Cambridge, Cambridge University Press, 2011, chap. 9 : « Climate Change », p. 216-239.
- ³⁰ Singer, *One World*.
- ³¹ Singer, Peter, *The Most Good you Can Do. How Effective Altruism is Changing Ideas About Living Ethically*, New Haven et Londres, Yale University Press, 2015.
- ³² Cette dernière critique est développée dans Renaut et coll., *Inégalités entre globalisation et particularisation*, p. 88-102.
- ³³ Birnbacher, Dieter, « What Motivates Us to Care for the (Distant) Future? », in Gosseries, Axel et Lukas H. Meyer, (dir.), *Intergenerational Justice*, Oxford, Oxford University Press, 2009, p. 273-300, p. 273-274. Voir également : Birnbacher, Dieter, « Some Moral Pragmatics of Climate Change », in Birnbacher, Dieter et Mat Thorseth, (dir.), *The Politics of Sustainability : Philosophical Perspectives*, New York et Londres, Routledge, 2015, p. 153-172.
- ³⁴ Sinäi, Agnès, « Introduction : L'Anthropocène, nouvelle catégorie de l'entendement », in Sinäi, Agnès, (dir.), *Penser la décroissance. Politiques de l'Anthropocène*, Paris, Presses de Sciences Po, 2013, p. 15-19, p. 17.
- ³⁵ Servigne, Pablo et Raphaël Stevens, *Comment tout peut s'effondrer. Petit manuel de collapsologie à l'usage des générations présentes*, Paris, Seuil, 2015.
- ³⁶ *Ibid.*, p. 115.
- ³⁷ Guzman, Andrew T., *Overheated: The Human Cost of Climate Change*, Oxford, Oxford University Press, 2013, chap. 6 : « Climate Change is Bad For Your Health ».
- ³⁸ *Ibid.*, p. 170.

- ³⁹ GIEC 2014, « Résumé à l'intention des décideurs », publié dans Field, C. B. et coll., (dir.), *Changements climatiques 2014 : Incidences, adaptation et vulnérabilité. Contribution du Groupe de travail II au cinquième Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat*, Genève, Organisation météorologique mondiale, 2014, p. 1-34, p. 20.
- ⁴⁰ Welzer, Harald, *Les guerres du climat. Pourquoi on tue au XXI^e siècle*, trad. L. Bernard, Paris, Gallimard, 2009, p. 295.
- ⁴¹ Roberts, Timmons J. et Bradley C. Parks, *A Climate of Injustice: Global Inequality, North-South Politics, and Climate Policy*, Cambridge et Londres, The MIT Press, 2007, p. 2. L'excellente analyse des négociations climatiques proposée par ces auteurs vient d'être mise à jour dans le volume suivant : Cipler, David et coll., *Power in a Warming World. The New Global Politics of Climate Change and the Remaking of Environmental Inequality*, Cambridge et Londres, The MIT Press, 2015.
- ⁴² Ce champ interdisciplinaire est constitué au croisement des sciences des systèmes complexes, des sciences de la Terre, de l'atmosphère et du climat et de l'écologie globale. Il considère la Terre comme un système complexe, qui va de son cœur jusqu'à la haute atmosphère, et dont les sous-systèmes (atmosphère, biosphère, lithosphère, hydrosphère, etc.) sont traversés et reliés entre eux par d'incessants flux de matières et d'énergie.
- ⁴³ Rockström, Johan, et coll., « Planetary Boundaries: Exploring the Safe Operating Space for Humanity », *Ecology and Society*, vol. 14, n° 32, 2009.
- ⁴⁴ Archer, David, *The Long Thaw. How Humans are Changing the Next 100,000 Years of Earth's Climate*, Princeton, Princeton University Press, 2009.
- ⁴⁵ GIEC 2013, « Résumé à l'intention des décideurs », in Stocker, T. F. et coll., (dir.), *Changements climatiques 2013 : Les éléments scientifiques. Contribution du Groupe de travail I au cinquième Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat*, Cambridge & New York, Cambridge University Press, 2013, p. 3-29, p. 26.
- ⁴⁶ Teyssède, Anne, « Boucles de rétroaction », in Bourg, Dominique et Alain Papaux, (dir.), *Dictionnaire de la pensée écologique*, Paris, PUF, 2015, p. 35-40.
- ⁴⁷ Lenton, Timothy M. et coll., « Tipping Elements in the Earth's Climate System », *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, n° 6, 2008, p. 1786-1793, p. 1792.
- ⁴⁸ Chung, Ryoa, « Théories idéale et non idéale », in Jeangène Vilmer et Chung, (dir.) *Éthique des relations internationales. Problématiques contemporaines*, p. 86.
- ⁴⁹ World Resources Institute. CAIT: Climate Data Explorer [en ligne], consulté le 3 juillet 2015, <http://cait.wri.org/>.
- ⁵⁰ Harris, Paul G., « Diplomacy, Responsibility and China's Climate Change Policy », in Harris, Paul G., (dir.), *China's Responsibility for Climate Change*, p. 1-21.
- ⁵¹ Ren, Yuan, « Under the Dome: Will This Film be China's Environmental Awakening? », *The Guardian*, 3 mai 2015.
- ⁵² Light, Andrew et Gwynne Traska, « A Responsible Path: Enhancing Action on Short-Lived Climate Pollutants », in Heyward, Clare et Dominic Roser, (dir.), *Climate Justice in a Non-ideal World*, Oxford, Oxford University Press, 2016, p. 169-188.
- ⁵³ Mills, Luke, « Global Trends in Clean Energy Investment » [en ligne], Bloomberg New Energy Finance, oct. 2015, consulté le 10 déc. 2015, <http://about.bnef.com/content/uploads/sites/4/2015/10/2015-10-08-Clean-Energy-Investment-Q3-2015-factpack.pdf>
- ⁵⁴ Stern, Nicolas, « Economic Development, Climate and Values: Making Policy », *Proceedings of the Royal Society B*, vol. 282, 2015, p. 1-9, p. 5.
- ⁵⁵ Pour une synthèse de la littérature sur les co-bénéfices liés aux politiques de réduction, voir IPCC 2014, Clarke, Leon et coll., (dir.), « Assessing Transformation Pathways », in Edenhofer, Ottmar R. et coll., (dir.), *Climate Change 2014 : Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge/New York, Cambridge University Press, 2014, p. 413-510, p. 468-477.

- ⁵⁶ Moellendorf, Darrel, « Can Dangerous Climate Change be Avoided? », *Global Justice: Theory Practice Rhetoric*, vol. 8, n° 2, 2015, p. 66-85, p. 82-84.
- ⁵⁷ Organisation mondiale de la santé, « Cancérogénicité de la consommation de viande rouge et de viande transformée » [en ligne], 2015, consulté le 10 déc. 2015, <http://www.who.int/features/qa/cancer-red-meat/fr/>
- ⁵⁸ Food and Agricultural Organization, « Livestock's Long Shadow. Environmental Issues and Options » [en ligne], Rome, 2006, consulté le 10 déc. 2015, <http://www.fao.org/docrep/010/a0701e/a0701e00.htm>
- ⁵⁹ Voir notamment Jeangène Vilmer, Jean-Baptiste, *Éthique animale*, Paris, PUF, 2008, chap. 8 : « Les animaux de consommation », p. 169-182.
- ⁶⁰ *Ibid.*, p. 174-175.
- ⁶¹ Broome, John, « Do not Ask for Morality », conférence donnée au Collège de France dans le cadre du colloque « Comment penser l'Anthropocène ? Anthropologues, philosophes et sociologues face au changement climatique », nov. 2015, consulté le 18 déc. 2015, <https://www.youtube.com/watch?v=8kl0Ur8xEVU>.
- ⁶² Posner, Eric A. et David Weisbach, *Climate Change Justice*, Princeton, Princeton University Press, 2010.

LA CONSERVATION AUTOLOGUE DE SANG DE CORDON OMBILICAL : VERS UNE NOUVELLE FORME DE PARTICIPATION BIOCITOYENNE ?

ANOUCK ALARY

CANDIDATE AU DOCTORAT EN SOCIOLOGIE, UNIVERSITÉ DE MONTRÉAL

RÉSUMÉ :

La transformation du sang placentaire en une précieuse source de cellules souches a donné naissance à partir des années 1990 à une industrie globale de conservation de sang de cordon ombilical faisant désormais concurrence à un large réseau de banques publiques de sang de cordon. Cet article explore les soubassements socioculturels liés à l'émergence de cette industrie et tente d'élucider les enjeux éthiques et politiques qu'elle pose. Si les banques publiques de sang de cordon sont porteuses des valeurs d'altruisme et de solidarité nationale traditionnellement liées au modèle redistributif d'échange de sang et d'organes né après la Seconde Guerre mondiale, les banques privées renvoient, elles, à des formes de solidarité bien différentes. C'est effectivement sous couvert de la solidarité familiale et de la responsabilité morale des mères de protéger leurs enfants qu'elles définissent la conservation privée comme une forme d'« assurance biologique » contre les risques à la santé de l'enfant. En permettant aux mères d'investir leurs tissus corporels à la fois dans le futur incertain de leurs enfants et dans des thérapies cellulaires expérimentales, ces banques promeuvent un nouveau modèle de participation du/de la patient.e à la coconstitution de futures innovations thérapeutiques. Nous inscrivons ce modèle de participation dans les reconfigurations contemporaines du biopolitique que le sociologue britannique Nikolas Rose (2007) voit s'incarner dans l'émergence d'une nouvelle forme de biocitoyenneté. L'article critique finalement ces services personnalisés en soulignant qu'ils ont le potentiel d'opérer de nouvelles formes de coercition sur les mères, dans un contexte sociopolitique caractérisé par une responsabilisation accrue des individus au regard de la « bonne gestion » de leurs risques à la santé. Ces services sont aussi jugés problématiques sur le plan éthique, au vu de leur incohérence avec un principe de justice distributive défendant l'accès égal pour chaque citoyen.ne à des soins de santé de base.

ABSTRACT:

Since the 1990s, the transformation of cord blood into a precious source of stem cells has given rise to a global cord blood bank industry, which is now competing with a large network of public cord blood banks. This article explores the sociocultural context surrounding the emergence of this industry and aims at elucidating the ethical and political concerns that this industry raises. Whereas public cord blood banks are purveyors of values such as altruism and national solidarity traditionally linked to the redistributive model of human blood and organ donation that emerged after World War Two, private banks engage very different forms of solidarity. It is indeed under the guise of family solidarity and mothers' moral responsibility to do their best to protect their children that they come to define private preservation as a form of "biological insurance" against potential risks to their children's health. By allowing mothers to invest their bodily tissues simulta-

neously in their children's uncertain future and in experimental cell therapies, these banks promote a new model of patient participation in the development of future therapeutic innovations. This article analyzes this model as embedded in the contemporary reconfiguration of biopolitics that sociologist Nikolas Rose (2007) envisions as constituting a new form of *biocitizenship*. The paper finally casts a critical eye on these personalized services, in particular because of the new forms of social coercion on mothers that are brought about in a social and political context characterized by an increasing moral accountability of individuals with regard to the "good management" of their own health risks. These services also appear ethically problematic in relation to political demands of distributive justice—e.g., the ideal of equal access to basic health care for every citizen.

1. INTRODUCTION : MARCHANDISATION DES ÉCHANGES DE TISSUS HUMAINS ET NOUVEAUX IMAGINAIRES DU CORPS (BIO-)POLITIQUE

Au lendemain de la Seconde Guerre mondiale, de nombreuses nations occidentales¹ ont soumis les échanges de tissus biologiques humains aux modalités d'une économie redistributive, telles que défendues par le politologue anglais R. Titmuss dans sa célèbre thèse *The Gift Relationship : From Human Blood to Social Policy* (Titmuss, 1970). Écrite dans la seconde moitié des années 1960, celle-ci représentait un plaidoyer pour le maintien d'un État-providence que Titmuss sentait menacé par les considérations économiques qui commençaient à gagner du terrain dans les cercles officiels du Parti travailliste (Fontaine, 2002). Afin de défendre un système de conservation de sang basé sur le don altruiste et anonyme, à l'époque encore dominant au Royaume-Uni, contre un système marchand dans lequel le sang était vendu et acheté, prévalent aux États-Unis, il a alors présenté les échanges allogéniques de tissus comme les garants de la « solidarité sociale ». Comprenant les produits du corps humain comme des biens étroitement liés à l'humain, il rejoignait le discours concernant la dignité incommensurable du corps humain, développé dans le contexte international à la suite des atrocités perpétrées durant la Seconde Guerre mondiale (Palvia, 1999) afin de l'opposer à l'idée d'un marché des corps perçu comme déshumanisant.

En parallèle à la mise en place progressive de ces modalités d'échange de tissus « par le don » dans la plupart des États industrialisés,² se sont toutefois consolidées depuis la fin du XX^e siècle, de manière globale, des pratiques d'exploitation du vivant fondées sur des logiques économiques et sociales radicalement différentes. La vitalité du corps humain s'est en effet trouvée projetée au cœur de formes inédites de marchandisation et de privatisation du biologique, qui ont d'ailleurs été définies comme parties prenantes d'un « second mouvement des enclosures »³ concernant le vivant (Boyle, 2003, p. 37). Ces nouvelles pratiques sont liées aux tendances majeures qui ont transformé l'organisation du secteur de la santé à partir du milieu des années 1980, au moment où les infrastructures cliniques et scientifiques de la médecine ont été restructurées autour de la science computationnelle et des innovations technoscientifiques⁴ (Clarke et coll., 2010). À la suite de ce « tournant technoscientifique », les thérapies et les méthodes de recherche et de diagnostic se sont en effet trouvées de plus en plus dépendantes de l'instrumentalisation d'organismes vivants d'origine humaine (Hoeyer, 2010). Afin de soutenir cette demande croissante, les pratiques consistant à extraire ou à récupérer certains tissus originaires du corps humain⁵ se sont multipliées et ces tissus ont été mis en circulation sur différents circuits d'échange dépassant largement les frontières et les régulations étatiques.⁶ Au sein de ces nouvelles « économies de tissus » (Waldby et Mitchell, 2007) globalisées, de nombreux États continuent toutefois d'organiser la récupération et les échanges de ces tissus biologiques selon le cadre légal « traditionnel » que de nombreux pays occidentaux avaient institutionnalisé après la guerre. Les banques allogéniques créées à la fin du XX^e siècle dans de nombreux pays industrialisés afin de soute-

nir la récupération et la manipulation de nouveaux biomatériaux humains destinés à être transplantés⁷ ont ainsi été fondées dans l'objectif d'assurer la collectivisation de ces produits au sein de systèmes de santé gratuits, universels et publics.⁸ C'est le cas des banques publiques de cellules souches allogéniques de sang de cordon, qui restreignent la manière dont les échantillons de sang de cordon peuvent être récupérés et exploités aux conditions du « don altruiste », c'est-à-dire au don anonyme de ces tissus entre les membres d'une même communauté nationale. À travers la manière dont ces banques conditionnent les pratiques de manipulation et les échanges du corps qu'elles opèrent, elles contiennent ainsi d'incorporer certains principes normatifs tels que les valeurs de solidarité sociale, d'altruisme et de réciprocité. Elles dessinent ainsi des manières d'appartenir à une communauté sociale constituée de formes d'obligations et de responsabilités bien précises. Ces formes d'appartenance s'inscrivent, comme nous l'exposerons, dans le prolongement du régime biopolitique⁹ qu'a constitué l'attribution par l'État d'un nouveau « droit à la santé » aux citoyen.nes, à travers l'instauration de l'État-providence au sein de nombreuses nations occidentales durant l'après-guerre (Foucault, 1994b).

À contre-pied des échanges opérés par les banques publiques de cellules souches allogéniques, un large secteur commercial de conservation de cellules souches de sang de cordon, dit « autologue »¹⁰ se développe depuis la fin des années 1990 à une échelle globale. Ces banques commerciales encouragent les nouvelles mères non pas à « donner » leurs tissus ombilicaux de manière anonyme aux autres membres de la communauté politique, mais à les préserver pour leurs propres enfants, au sein d'un compte de conservation payant et exclusif. Bien que l'efficacité thérapeutique des cellules souches autologues n'ait pas encore été démontrée, l'idée promulguée par ces banques est que leur congélation permettrait aux mères d'en collecter les bienfaits futurs sous forme de thérapies cellulaires autologues dites « en développement ». En offrant aux mères de constituer une partie de leurs propres corps comme une forme d'« assurance biologique » contre la maladie et le vieillissement cellulaire, ce secteur autologue paraît alors être fermement ancré dans le domaine des nombreuses promesses entourant le secteur de la médecine régénérative. Le secteur commercial de conservation du sang de cordon propose une approche tout à fait innovatrice afin d'articuler la relation entre les individus et leurs tissus biologiques, fondée notamment sur de nouvelles formes de possession et d'investissement de soi dans le futur des technosciences. Il s'établit ainsi radicalement à rebours de l'économie traditionnelle du don « altruiste », engagé à soigner des personnes présentement malades.

La coexistence de banques publiques et de banques commerciales dans la plupart des pays industrialisés génère de ce fait une controverse examinée dans des avis publiés par les comités consultatifs d'éthique de plusieurs pays¹¹ et par diverses associations professionnelles de santé.¹² Le présent article vise à comprendre et à expliquer, dans une perspective sociologique, les tensions et clivages naissants de la coexistence actuelle des banques publiques et privées de sang de cordon. Il se propose ainsi de donner un éclairage nouveau sur les enjeux éthiques et

sociaux soulevés par les banques privées de sang de cordon en identifiant et en comparant les formes de solidarité¹³ que les circuits allogéniques et autologues sous-tendent respectivement. Pour ce faire, nous procéderons tout d'abord à une analyse de la manière dont les banques allogéniques et les banques autologues définissent la matérialité du sang de cordon (son potentiel et sa valeur), la santé et le rôle des patient.es, ainsi que l'imbrication de ces conceptions avec l'imaginaire plus large de la citoyenneté et de la communauté politique. Nous analyserons les formes de solidarité construites par les banques autologues comme une conséquence de la relation de propriété que les banques autologues établissent entre les mères et leurs tissus biologiques, permettant à celles-ci d'*investir* conjointement dans le futur de leurs enfants et celui de la médecine. Nous défendrons l'hypothèse selon laquelle ce nouveau modèle de participation des patient.es à la coconstruction de nouvelles thérapies biomédicales s'inscrit dans les reconfigurations contemporaines du rapport à la santé que le sociologue Nicolas Rose voit s'incarner dans l'émergence d'une nouvelle « citoyenneté biologique » (Rose, 2007). Nous discuterons ensuite des normativités entourant la gestion prudente par les individus de leur santé que ce modèle de participation valorise, cela dans un contexte sociopolitique marqué par une recodification des défis de santé en problèmes de *management* personnel et par une culpabilisation des individus liée à leur échec à rester en santé. Nous montrerons également que ce modèle a le potentiel de générer de nouvelles formes d'exclusion sociale pour les individus et les groupes n'ayant pas les capacités financières leur permettant d'investir dans leur propre « capital biologique ». En illustrant la possibilité pour les citoyen.nes détenant le capital financier nécessaire afin de conserver leurs tissus dans un compte familial de se dissocier des responsabilités qui pendant longtemps allaient de pair avec le droit, assuré par l'État-providence, à des soins de santé publics et universels, ce secteur nous paraît opérer une stratification des vies et des corps corrélée aux inégalités socioéconomiques existant au sein des populations locales.

2. QUE SONT LES CELLULES SOUCHES DE SANG DE CORDON ?

Les cellules souches sont des cellules encore totalement ou en partie indifférenciées, alors capables de se multiplier à l'infini et de se spécialiser en différents types de cellules. Par ces capacités, elles semblent pouvoir générer une véritable révolution biomédicale dont le paradigme serait celui d'une nouvelle médecine régénérative se proposant d'optimiser les mécanismes impliqués dans l'autogénération des cellules dans l'objectif de développer des thérapies cellulaires afin de réparer des tissus et organes endommagés ou vieillissants. Pouvant se résumer à l'idée « d'aider le corps à se soigner lui-même » (*European Technology Platform on Nanomedicine*, 2005, p. 27, cité dans Lafontaine et Noury, 2014, p. 34), la médecine régénérative promet de permettre un jour de contourner les difficultés liées à l'obtention d'organes ou de parties d'organes compatibles avec tou.tes les patient.es, ainsi que les dommages liés à leur extraction lorsque celle-ci est réalisée sur des personnes vivantes. Si certaines thérapies régénératives sont présentement pratiquées, notamment les greffes autologues de sang de moelle osseuse ou de sang périphérique en cancérologie, la plupart d'en-

tre elles se situent à ce jour à un stade très expérimental (Waldby, 2006, p. 56). Bien que les cellules souches embryonnaires, parce qu'elles sont encore entièrement indifférenciées, soient les seules à pouvoir se convertir en n'importe quel type cellulaire, enflammant ainsi l'imagination des scientifiques, la charge symbolique attribuée aux embryons limite leur utilisation en tant que simples sujets d'expérimentation ou instruments biomédicaux. C'est pourquoi les chercheuses et chercheurs se sont tourné.es à partir du début des années 1970 vers des sources de cellules plus facilement récupérables et manipulables telles que le sang périphérique et de moelle osseuse, ou encore certains tissus provenant du corps des femmes, comme le sang menstruel ou placentaire. La découverte de larges colonies de cellules souches hématopoïétiques dans le sang placentaire à partir du milieu des années 1970 fut ainsi largement tributaire du fait que les cordons ombilicaux et les placentas, longtemps définis comme des « déchets opératoires »,¹⁴ purent être récupérés et mobilisés sans grande contrainte au sein de divers projets profitables de « recyclage » (Waldby et Mitchell, 2006, p. 115).

2.1. Les usages thérapeutiques présents des cellules souches de sang de cordon en oncologie

Les cellules souches hématopoïétiques étant à l'origine des différentes lignées sanguines de l'organisme, les cellules souches de sang de cordon ont rapidement été évaluées comme de potentiels outils thérapeutiques, pouvant permettre de contourner certaines des limites posées par les greffes de cellules souches hématopoïétiques issues de la moelle osseuse. Les cellules souches de moelle osseuse étaient déjà utilisées depuis la fin des années 1960 afin de soigner les patient.es atteint.es de cancers du sang, notamment de la leucémie. La transplantation de cellules souches provenant d'un tissu néonatal tel que le sang placentaire, du fait de l'immaturation du système immunitaire à la naissance, n'exige toutefois pas de compatibilité immunologique complète entre donneuse et receveuse/receveur (Kline, 2001), comme cela est le cas pour les greffes de cellules souches de moelle osseuse. Elle permet alors de multiplier le nombre de donneuses compatibles pour un.e même patient.e en dehors des membres de sa propre famille. D'autre part, les échantillons de sang de moelle osseuse ne peuvent être extraits qu'à travers des procédures lourdes et invasives impliquant l'hospitalisation des donneuses/donneurs ainsi que leur maintien sous anesthésie générale ou spinale. Ils ne sont donc pas collectionnés de façon routinière au sein de banques thérapeutiques, mais simplement emmagasinés de manière virtuelle dans des « banques de données », soit des registres informatisés où se trouvent consignées les informations concernant des donneuses/donneurs potentiels. La collecte du sang de cordon, dite (à tort) par la plupart des bioéthiciens et des chercheuses et chercheurs, pratiquée *après* la naissance de l'enfant,¹⁵ est appréhendée de son côté comme non invasive pour la donneuse. Elle permet de ce fait le développement de banques constituées d'unités de sang congelées, c'est-à-dire « réelles » et directement disponibles. Toutefois, compte tenu du faible volume de sang pouvant être recueilli, l'usage thérapeutique des cellules souches de sang placentaire reste à ce jour réservé à des enfants de moins d'une vingtaine de kilos (Waldby, 2006).

2.2. Les cellules souches de sang de cordon, des cellules pour la médecine régénérative du futur

Le succès thérapeutique du sang de cordon en oncologie et le fait que son extraction ne semble poser « aucune préoccupation éthique »¹⁶ semblent aussi avoir facilité son insertion au sein de multiples recherches visant à instrumentaliser son potentiel régénératif dans le développement de futures thérapies cellulaires. Ces recherches sont elles-mêmes soutenues financièrement par les investissements massifs de puissantes entreprises privées, en percevant déjà les grandes retombées économiques (Boileau, 2002). Présentement, il existe ainsi deux voies principales au sein de ces recherches. La première se situe dans le développement de méthodes permettant d'étendre la masse de cellules souches de sang de cordon *in vitro* afin que de plus larges volumes puissent être transplantés, multipliant ainsi le nombre de potentiels receveurs et receveuses de greffes au-delà de jeunes enfants (Egan, 2000). L'autre est orientée vers le développement de stratégies permettant de transformer les cellules souches de sang de cordon en cellules pluripotentes induites, c'est-à-dire de manipuler leur temporalité afin de les inciter à retourner à leur passé embryonnaire et les contraindre ainsi à se spécialiser en n'importe quel type de tissu (Waldby, 2006). C'est précisément là que ces cellules paraissent les plus séduisantes pour les chercheuses et chercheurs en médecine régénérative. Envisagé comme un moyen d'éviter les enjeux éthiques liés à l'utilisation biomédicale de cellules souches embryonnaires, le sang de cordon est présentement appréhendé comme n'étant rien de moins que le « futur de la médecine régénérative » (Kedereit et Rudolph, 2010). Ce secteur médical relativement nouveau, couvrant des recherches assez hétéroclites, se concentre sur l'isolation et l'exploitation de la capacité des cellules souches à se reproduire et à se spécialiser en différents types de cellules. L'objectif visé est d'aider ainsi le corps à dépasser ses propres limites – notamment la dégénération cellulaire liée au vieillissement. Il constitue un domaine particulièrement emblématique de l'importance croissante des investissements spéculatifs dans la recherche, orientant celle-ci vers les développements très prometteurs de futures thérapies cellulaires. Au-delà de ses applications présentes en oncologie, le sang de cordon se voit alors rapidement attribuer, par un ensemble de firmes commerciales, un potentiel dans le développement de futures thérapies cellulaires autologues. Ces banques affirment que de telles thérapies personnalisées permettront un jour de soigner de nombreuses maladies dégénératives autant chez les enfants que chez les adultes (telles que la sclérose en plaques, les maladies de Pick, d'Alzheimer, de Huntington, de Parkinson, etc.), d'éradiquer les signes du vieillissement ou même de prolonger la vie.

3. LES BIOBANQUES : LE CORPS MATERNEL «BIO-OBJECTIVÉ» ET DISTRIBUÉ DANS LE TEMPS ET DANS L'ESPACE

Les nouvelles mères peuvent désormais, avant la naissance de leur enfant, choisir de faire don d'un échantillon de leurs cellules à des banques publiques, développées afin de soutenir la généralisation des greffes allogéniques de sang de cordon en oncologie clinique, ou de conserver celui-ci au sein de banques privées

exclusivement familiales, fondées sur les promesses multiples dont sont investies les cellules souches de sang de cordon concernant le développement potentiel de nouvelles thérapies cellulaires autologues. Afin de rendre les cellules reproductives du corps maternel manipulables et échangeables au sein des circuits de régénération allogéniques et autologues, les banques publiques et privées s'appuient respectivement sur la cryoconservation, un procédé technique permettant de congeler des tissus ou des cellules à des températures extrêmement basses pour une utilisation ultérieure. Ce processus, à travers lequel les cellules de sang de cordon ombilical seront « bio-objectivées », ¹⁷ c'est-à-dire stabilisées sous une forme matérielle, transformées en biomédicaments, en « choses fabriquées », représente effectivement le passage obligé de leur mise en circulation sur différents circuits d'échanges. En manipulant artificiellement le temps et l'espace biologique dans lequel ces cellules évoluent, ces banques les maintiennent en effet dans un espace liminal entre la vie (*l'animation cellulaire*) et la mort (la *cessation cellulaire*) pour des durées qui transgressent la temporalité naturelle de la dégénération cellulaire (ces cellules peuvent être préservées ainsi pendant une vingtaine d'années). La soumission des cellules de sang de cordon à cette temporalité particulière, communément définie en biologie cellulaire comme un état de « suspension » (Franklin et Lock, 2003) permet alors de détacher ces cellules de la « normativité d'un ordre vital donné » (Rose, 2007, p. 14). Elle les rend autonomes quant aux contraintes naturelles posées par le cycle du corps maternel dont elles proviennent, afin de pouvoir les préserver dans le temps et de les transporter dans l'espace, vers d'autres corps. Métamorphosées en entités biologiques aux frontières fluides, elles peuvent être soumises aux intentions humaines de façon virtuellement infinie et autoriser alors leur entrée sur de multiples circuits d'échanges de tissus humains.

Cet état « suspendu » suppose toutefois une redéfinition anthropologique du vivant en tant que phénomène technologique et essentiellement plastique (Landecker, 2007). ¹⁸ La naturalisation de cette conception *technicisée* du vivant, en tant que substance que l'on peut réinventer et remodeler à souhait, a des implications quant à la manière dont l'identité humaine est elle-même culturellement appréhendée, redéfinissant aujourd'hui pour les êtres humains « la tâche sociale et culturelle [...] d'être simultanément des choses biologiques et des personnes humaines » (*ibid.*, p. 235). À travers la cryoconservation des cellules souches de sang de cordon, c'est en effet le corps reproductif féminin qui se voit projeté au-delà des catégories conventionnelles du vivant (telles que les catégories de nature et de culture, de vie et de mort, d'êtres animés et d'objets), soit imaginé comme un domaine de pure potentialité. Ces cellules sont intégrées à travers l'intermédiaire des banques publiques et privées de sang de cordon, au sein de différentes économies de tissus interrogeant « la fabrication et l'essence de l'être humain, tant dans sa matérialité biologique que dans sa dimension anthropologique » (Boileau, 2002, p. 21), à travers la façon dont ces banques réorganisent les frontières rendues flexibles du corps maternel. ¹⁹ Ces deux types de banques interrogent ensuite la question du *lien social*. En négociant très différemment le statut fluide et non déterminé des cellules souches immortalisées de sang de cordon, et en définissant de manière contradictoire l'autorité que les individus sont en

mesure d'exercer sur celles-ci, ces institutions génèrent des imaginaires du corps et de la responsabilité des mères tout à fait distincts. Ces imaginaires sociaux renvoient eux-mêmes, comme nous l'exposerons, à deux régimes biopolitiques tout aussi distincts : 1) à une biopolitique redistributive et assistantielle caractérisée par une mutualisation des risques sociaux, incluant ceux liés à la santé; et 2) à une biopolitique néolibérale se définissant par une individualisation des risques, et impliquant une responsabilisation morale des citoyens quant à la prise en charge prudente de leur propre santé et celle de leurs proches.

3.1. Les banques publiques de sang de cordon : parties prenantes d'une économie redistributive

Les premières banques publiques de sang placentaire apparaissent dans la plupart des pays de l'Organisation de coopération et de développement économique (OCDE) quelques années après la première transplantation allogénique réussie de sang de cordon. Cela marque la naissance de ce qui deviendra peu à peu un réseau transnational de banques publiques, existant dans plus de 30 pays et desservant près de 500 centres de transplantation autour du monde. Dans de nombreux pays industrialisés d'Amérique du Nord et Centrale, d'Europe et d'Asie, certaines mères se voient alors invitées à faire don d'un échantillon du sang de cordon à des banques publiques reposant sur les retombées *présentes* de cellules souches en oncologie. Ces échantillons seront effectivement accumulés au sein de ces banques afin d'assurer l'accès à des thérapies cellulaires allogéniques à des enfants leucémiques. Les cellules souches de sang de cordon, bien qu'elles représentent au sein de ces banques des outils cliniques aux retombées thérapeutiques avérées, ne semblent toutefois pas définies par celles-ci comme de simples produits pharmaceutiques, des objets « ordinaires » d'échange marchand. Elles sont, en étant définies comme des dons sans réserve – altruistes, volontaires, anonymes et gratuits –, réintégrées dans les formes de circulation symbolique dessinées par l'« économie du don » qui a gouverné la plupart des systèmes de sang et d'organes nationaux au courant de la seconde moitié du XX^e siècle. Comme l'illustrent bien les propos tenus par le Groupe européen d'éthique des sciences et des nouvelles technologies (GEE) dans un avis concernant les aspects éthiques de la conservation du sang de cordon ombilical (GEE, 2004), la donation d'un échantillon de sang placentaire au secteur de conservation public est en effet généralement définie comme « un rare et louable exemple d'altruisme », un acte qui « contribue à la cohésion sociale » impliquant « une solidarité ou une générosité » (Annas, 1999, p. 1522). La redistribution sur des réseaux d'échanges allogéniques de tissus corporels humains telle qu'opérée par les institutions biomédicales publiques a en effet traditionnellement *incarné* le maintien, la reproduction de formes spécifiques de solidarité. Développées à Barcelone durant la Guerre civile espagnole dans l'effort de guerre mobilisé afin d'approvisionner les soldats en unités de sang et en organes (Rabinow, 1999), les méthodes de conservation et de transplantation sanguine sont historiquement liées à un imaginaire patriotique qui résonne lui-même avec des associations plus anciennes liant le sang à la race, et la race à la citoyenneté nationale (Foucault, 1980). Lorsqu'elles furent perfectionnées aux États-Unis, au

Royaume-Uni et en Afrique du Nord après la Seconde Guerre mondiale, elles continuèrent « de porter sur elle la marque de la solidarité nationale, d'un geste volontaire et bénévole, de l'effort collectif de la nation entière » (Rabinow, 1999). La période des Trente Glorieuses se caractérise en effet par la consolidation par la majorité des États occidentaux d'un État-providence, soit de dispositifs visant à « maintenir ou reconstituer la société dans son unité relative, contre les multiples possibilités d'éclatement que porte sa modernisation, pour compenser les désavantages les plus criants, réparer les dysfonctionnements les plus évidents, garantir un minimum de sécurité à tous » (Castel, 1988, p. 68). Dans ce contexte de grandes mutations sociales et politiques, où l'État se met à attribuer certains droits sociaux fondamentaux aux individus afin de leur assurer une protection minimale, on aurait assisté à l'extension du pacte *biopolitique*.²⁰ La mise en place de systèmes de santé publics aurait marqué l'émergence d'une véritable « somatocratie »,²¹ propulsée par les finalités que deviennent l'hygiène et la santé corporelles des citoyen.nes. Le plan Beveridge développé en 1942 en Grande-Bretagne, qui a par la suite servi de modèle pour l'Organisation de la santé, a alors représenté la première véritable insertion de la santé dans le champ des grandes dépenses économiques de l'État. Il participait précisément d'un projet de redistribution économique : garantir à tou.tes les mêmes chances de se soigner afin de corriger l'inégalité des revenus (Foucault, 1974, p. 42). Inscrites dans cette volonté de l'État social d'assurer à chaque citoyen.ne un accès égal à des soins de santé, les banques publiques de sang et d'organes se sont alors caractérisées par le rejet d'un modèle marchand dans lequel le sang et les organes seraient vendus et achetés librement, en les définissant comme des « éléments » devant être « versés au patrimoine commun » (Langlois, 1998, p. 80), offerts gratuitement par les citoyen.nes aux autres membres de la communauté nationale. La célèbre thèse de R. Titmuss (1970), ayant par la suite largement influencé la plupart des politiques sociales des États industrialisés concernant les modalités d'échange à instaurer, révèle particulièrement bien en quoi la soumission des échanges de tissus biologiques aux conditions du « don altruiste » entre citoyen.nes participait de la cristallisation volontaire de liens sociaux et politiques spécifiques. En opérant une séparation stricte entre la forme du « don », liée à la sphère des relations sociales, de la moralité et de la cohésion sociale, à celle de la « marchandise », appartenant au domaine impersonnel, rationnel et calculateur de l'économie, il reprenait de manière explicite l'argument développé par Marcel Mauss (2012 [1924]). Ce dernier avançait en effet que les échanges de « dons » étaient, dans les sociétés traditionnelles, au fondement de liens d'obligation entre les personnes en impliquant nécessairement l'endettement symbolique des réceptrices et récepteurs envers les donneuses et donneurs, soit une réciprocité, un « contre-don ». L'échange anonyme et gratuit de sang entre inconnu.es avait ainsi pour Titmuss l'effet de réaffirmer les valeurs civiques au fondement de la communauté politique. Il permettrait d'assurer, à travers une confiance mutuelle, la permanence d'un lien social basé sur la « socialité secondaire » (Godbout et Caillé, 1992), impersonnelle de principe, existant entre les citoyen.nes, au fondement de la « communauté imaginée de la nation » (Anderson, 1983). Cette « exaltation pour le don » paraît ainsi rassembler encore aujourd'hui les institutions biomédicales publiques autour de « ce qu'elles voient

sans métaphore comme une redistribution de l'énergie vitale commune à tout le corps social » (Moulin, 2005, p. 51).

Si au sein des échanges allogéniques de sang de cordon, les capacités génératives propres à la biologie reproductive du corps des femmes semblent alors se voir mobilisées non plus vers la création de nouveaux êtres, mais vers la *régénération* du corps social, « prêtes à revitaliser le corps postnatal malade ou vieillissant » (Waldby, 2006, p. 61) de la nation, qu'en est-il des banques privées? Quelles formes de liens sociaux et d'engagement *par* le corps valorisent-elles de leur côté et en quoi ceux-ci divergent-ils des relations de solidarité sociale incarnées par le secteur public de sang de cordon?

3.2. Les banques autologues de sang de cordon et la privatisation du don « solidaire » en tant que bio-investissement familial

Tandis que Byocyte Corporation, la première banque autologue de sang de cordon est née aux États-Unis en 1993 (Holden, 1993) – soit l'année où les premières banques de sang de cordon sont également apparues dans le secteur public –, des banques commerciales sont aujourd'hui présentes dans la plupart des nations industrialisées. Il était ainsi déjà possible, à la fin de l'année 2014, de comptabiliser 207 banques privées dans plus de 54 pays, détenant une réserve de plus de 4 millions d'unités de sang en 2014, contre seulement 158 banques publiques présentes dans 36 pays, détenant moins de 731 000 unités de sang de cordon (World Marrow Donor Association, 2014). Ces banques proposent aux nouvelles mères de conserver un échantillon de sang de cordon, récolté au moment de la mise au monde de leur nouveau-né dans un compte privé réservé à un usage strictement familial. Définissant leur service comme une forme de « bio-assurance » contre les maladies potentielles de l'enfant à naître ou de sa parenté proche,²² elles incitent les mères, à travers une publicité massive, à constituer pour leur famille une réserve de cellules souches disponible en tout temps. Pour cela, les mères sont invitées à réaliser un investissement initial allant de 900 à 2 300 dollars incluant le transport et l'entreposage la première année, puis de déboursier une centaine de dollars chaque année afin de préserver l'unité de sang congelée.²³ Selon les banques privées, cet échantillon pourrait tout d'abord être utilisé en tant que greffon dans le cas où l'enfant lui-même développerait, et cela avant d'atteindre la vingtaine,²⁴ une maladie pouvant être soignée grâce à une greffe « autologue » de cellules souches. Or, l'application présente des greffes de cellules souches de sang de cordon est à ce jour presque exclusivement réalisée à partir de cellules allogéniques.²⁵ Ces banques privées s'appuient alors avant tout sur des développements *potentiels* de nouvelles thérapies autologues à base de cellules souches. Elles s'instituent ainsi à rebours des principes de la médecine fondée sur les données probantes (*evidence-based medicine*). Les banques privées affirment également que la conservation de ces cellules dans un compte familial pourrait être utile dans le cas où l'un.e des frères et sœurs du nouveau-né développerait une maladie pouvant être soignée par une transplantation allogénique de cellules souches. Or, la plupart des banques publiques de sang de cordon offrent elles aussi la possibilité de conserver celles-

ci à titre familial, et cela gratuitement, dans le cas peu probable où un membre de la parenté génétique du nouveau-né développerait effectivement une maladie rare (Dickenson, 2013, p. 92). Devant l'absence de résultats cliniques soutenant les greffes de cellules souches de sang de cordon autologues, de nombreuses associations de professionnels de la santé, notamment l'American Academy of Pediatrics (AAP), The American College of Obstreticians and Gynecologists (ACOG), la World Marrow Donor Association (WMDA) et, au Canada, The Society of Obstetricians and Gyneacologists Canada (SOGC) ont de concert déploré la logique hautement spéculative de ce service.²⁶ Les pratiques de ces banques ont aussi fait l'objet de la condamnation sans équivoque de certains États européens tels que la France, l'Espagne, la Suisse, le Luxembourg, la Belgique et l'Italie, ayant interdit l'implantation de tels services sur leurs territoires respectifs. De nombreux comités consultatifs d'éthique aux échelles européenne et nationale dénoncent également dans leurs avis respectifs le caractère largement spéculatif des promesses thérapeutiques des cellules de sang de cordon sur lesquelles ces banques privées s'appuient.²⁷ Ils déplorent aussi leur logique commerciale, qui leur semble inévitablement aller à l'encontre d'un système de santé public et universel et de l'intérêt collectif. Cela est bien exemplifié par les propos tenus en 2002 par le Comité consultatif national d'éthique pour les sciences de la vie et de la santé (CCNE) en France :

La conservation du sang placentaire pour l'enfant lui-même apparaît comme une destination solitaire et restrictive en regard de la pratique solidaire du don. Il s'agit d'une mise en banque de précaution, d'une capitalisation biologique préventive, d'une assurance biologique dont l'utilité effective, dans l'état actuel de la science apparaît bien modeste. (CCNE, 2002, p. 9)²⁸

En paraissant capitaliser massivement, au nom d'un principe de précaution douteux, sur ce qui pourrait autrement être rendu gratuitement disponible aux besoins présents du public dans les pays détenant déjà un système allogénique et solidaire d'échange de sang,²⁹ ce secteur est alors accusé de priver certains enfants malades de ce qui est souvent la meilleure thérapie possible pour les guérir, et d'ailleurs parfois la seule. Il semblerait toutefois que ce secteur ne soit pas uniquement condamné en raison des conséquences matérielles que représenterait une *perte* d'échantillons pour le secteur public. Il est effectivement aussi critiqué en raison de l'affront que l'existence de ce secteur commercial représenterait quant aux valeurs de solidarité sociale, d'altruisme et de générosité entre citoyen.nes portées par les échanges redistributifs de tissus biologiques humains, et les conséquences de celui-ci sur la précarisation du lien social et politique, sur les conditions du « vivre ensemble ». Tel que le souligne encore une fois de manière tout à fait explicite le CCNE, « le danger le plus grave est pour la société dans la mesure où l'instauration de telles banques est de nature à s'opposer au principe de solidarité, sans lequel il n'y a pas de survie possible pour une société, quelle qu'elle soit » (CCNE, 2002, p. 12, souligné dans le texte). Menaçant de décomposer la solidarité sociale incitée par les institutions de l'État social et de renvoyer les individus à eux-mêmes, le développement

croissant de banques autologues et commerciales de sang de cordon semble dès lors mettre en péril « l'identification traditionnelle entre l'État, le corps social et la solidarité nationale » (Santoro, 2009, p. 6) ayant longtemps été au fondement des échanges publics de tissus humains, soit toucher au fondement même du lien politique.

Les banques privées de sang de cordon paraissant ainsi entrer largement en conflit avec les principes portés par le secteur public de sang de cordon, comment saisir l'architecture morale spécifique qu'elles dessinent à leur tour, les valeurs et les responsabilités morales qu'elles portent? Nous tenterons dès à présent de rendre compte de celles-ci afin de mieux saisir ce que révèle le succès de ces banques autologues quant aux mutations des conceptions que nous nous faisons à la fois du corps, de la personne et du lien social dans les sociétés occidentales contemporaines.

4. LE PASSAGE DU « DON SOLIDAIRE » À UNE FORME DE BIO-INVESTISSEMENT FAMILIAL : UNE OUVERTURE SUR DES RECONFIGURATIONS CONTEMPORAINES DU BIOPOLITIQUE

Les partisan.es des banques publiques de sang de cordon expliquent généralement l'attrait des banques autologues comme le résultat des publicités mensongères qu'elles font largement circuler quant aux exploits des cellules souches autologues, se jouant de la naïveté des mères (Waldby, 2005, p. 15). Nous voudrions proposer que l'attrait de ce service ne repose pas uniquement sur l'usage promotionnel du mensonge, mais se situe également dans la manière tout à fait innovante dont ce secteur articule la relation entre les mères et leurs tissus biologiques dans un contexte caractérisé par l'incertitude (cela à la fois au regard des avancées de la médecine régénératrice et des risques à la santé de leurs propres enfants). En permettant aux mères de préserver un échantillon de sang de cordon pour l'usage exclusif de leurs propres enfants, les banques commerciales de sang de cordon inscrivent en effet la forme du « don » dans la continuité du lien affectif et obligationnel liant la mère à son enfant. En tension avec la solidarité nationale valorisée par les banques publiques de sang de cordon, elles valorisent la solidarité familiale tout en présentant la conservation comme le vecteur de son maintien. Le futur étant défini comme incertain au regard des risques à la santé de l'enfant et des possibilités thérapeutiques à venir, la conservation autologue semble en effet représenter la preuve tangible que la mère aura effectivement fait tout son possible afin de protéger son enfant. Si elle est définie comme une forme d'« assurance biologique » pour l'enfant, elle est également présentée comme une « assurance de conscience » pour la mère, un outil qui lui permettrait d'éviter d'expérimenter le lourd poids de la culpabilité dans le cas où le pire adviendrait sans qu'elle ait, au préalable, pris les mesures appropriées.

4.1. L'investissement des mères dans le potentiel thérapeutique attribué au sang de cordon et la construction du nouveau-né comme « malade présymptomatique »

Profitant de l'indétermination à laquelle le droit de propriété est relégué lorsqu'il doit se confronter à des parties biologiques *détachées* du corps,³⁰ les banques privées remettent en question la condition du « don altruiste » sans pour autant permettre aux femmes de faire directement commerce des produits de leur corps.³¹ Entre la mise en circulation de leurs propres cellules sous la forme de « dons altruistes » ou sous celle de simples objets de propriété pouvant être vendus sur le marché comme n'importe quelle marchandise, les mères se voient en effet offrir une troisième possibilité : celle de les constituer en une forme d'« investissement » biologique pour leurs enfants (Waldby et Mitchell, 2006, p. 123). La relation entre la mère et le sang de cordon s'appuie ici sur une forme nouvelle de contractualisme marchand permettant aux mères de préserver une *certaine* autorité sur le potentiel thérapeutique attribué à leurs cellules en les maintenant au sein d'un compte exclusivement familial. Bien que légalement, le droit de propriété sur le sang de cordon soit attribué à l'enfant sur la base d'une même identité génétique,³² et que ce droit semble être, au sein de ces banques, de plus conditionnel à l'obligation de payer la banque tous les mois,³³ il reste que la mère s'y trouve définie comme la « tutrice » du sang de cordon. L'enfant ne se verra en effet octroyer des droits sur « ses propres » cellules que lorsqu'il atteindra l'âge de la majorité, devenant ainsi une personne légale. Les mères, si elles ne peuvent pas tout à fait *posséder* les cellules provenant de leur propre corps à travers ces arrangements contractuels, se voient donc néanmoins offrir la possibilité d'investir stratégiquement dans leurs propres tissus, identifiés comme de futures ressources thérapeutiques sur la base des nombreuses promesses entourant le secteur de la médecine régénérative. La valeur du sang de cordon pour celles-ci n'est donc pas fixe; elle ne repose pas sur ses pouvoirs de régénération *présents*, mais sur « des possibilités futures de croissance » (Cooper, 2008, p. 8), soit sur une logique purement spéculative. Offrant aux mères d'isoler par cryoconservation un de leurs tissus corporels hors de la temporalité biologique naturelle de leurs propres corps pour une durée potentiellement infinie, ce service leur permet de constituer, à partir de leur corporalité, une réserve de cellules constamment à disposition. Prête à revitaliser à tout moment le corps de leur enfant qui, de son côté, continuera à vivre et à vieillir, ce morceau de corps « immortalisé » représente une potentielle esquivance à la sentence intolérable que semble constituer pour ces mères le destin biologique de leur enfant. Conceptualisant l'enfant comme un malade en devenir, ce service participe d'un effacement des frontières entre la santé et la maladie, entre le normal et le pathologique (Canguilhem, 1979 [1966]), au profit d'une logique préventive tout à fait douteuse. À travers la conservation du sang de cordon, il ne s'agit pas en effet pour la mère de permettre à son enfant déjà malade d'être soigné, ni même encore de souscrire à une stratégie préventive avérée, mais de se projeter mentalement dans un futur où son enfant serait non seulement atteint d'une maladie nécessitant une greffe de cellules souches, mais devrait de plus obtenir une greffe de « ses propres » cellules. Cette logique doublement spéculative

lative implique toutefois comme on l'a vu une intervention tout à fait tangible sur les corps du nouveau-né et de sa mère, ainsi que des risques tout aussi concrets sur leur santé respective.³⁴ Il nous semble que ces pratiques traduisent alors de manière tout à fait concrète l'évolution contemporaine du rapport entre médecine et société que tentent de saisir la sociologue A. Clarke et ses collègues en proposant de reconceptualiser la *médicalisation* (Zola, 1972) sous le terme de *biomédicalisation* (Clarke et coll., 2010). La médicalisation est un concept sociologique communément définie comme « un processus par lequel des problèmes non médicaux sont redéfinis et traités comme des problèmes médicaux, généralement en termes de maladies ou de troubles » (Gabe, 2013, p. 49). Au courant du XX^e siècle, de nombreux phénomènes sociaux jugés déviants ou potentiellement problématiques, tels que la naissance, l'homosexualité, la délinquance, le vieillissement, la mort et le deuil, ont en effet été redéfinis en problèmes médicaux, justifiant une intervention (pouvant être psychologique, psychiatrique, médicamenteuse, chirurgicale, etc.) relevant de l'autorité du médecin. La médicalisation se caractérisait selon Clarke et ses collègues par l'exercice par la médecine d'une force contraignante de *contrôle* sur les comportements et les corps des patient.es visant à les normaliser, favorisant alors l'expansion de l'autorité des médecins sur des pans toujours plus larges de la vie sociale (Clarke et coll., 2010, p. 51). Or, selon les auteures, si cette logique normalisante a été dominante depuis les années 1940 jusqu'aux années 1990, une autre logique serait venue prolonger celle-ci depuis le milieu des années 1980 (*ibid.*, p. 88). Le terme de biomédicalisation permettrait alors de capturer les mécanismes par lesquels la médecine, en devenant dépendante des innovations technoscientifiques (telles que les biotechnologies, la bio-ingénierie, la génomique, les thérapies géniques, la médecine régénérative, le clonage, etc.), se serait progressivement inscrite dans la voie d'une *transformation* des corps à travers des interventions technoscientifiques ayant lieu au cœur même des processus du vivant (*ibid.* p. 55). La médecine aurait adopté un nouveau rôle ne visant plus à ramener les comportements et les corps des patient.es vers la norme, mais plutôt à particulariser et à individualiser les corps en leur attribuant de nouvelles propriétés biologiques, soit à les transformer à l'échelle moléculaire en faisant usage de possibilités technoscientifiques croissantes. Si cette voie vers la transformation des corps est, comme le soutiennent ces auteures, effectivement en partie portée par des capacités technoscientifiques toujours plus importantes, elle nous semble également avoir été encouragée par la déconstruction au milieu du XX^e siècle des pôles de la santé et de la maladie en termes de *risque* (abaissant les seuils de la maladie en multipliant les états de santé jugés problématiques).³⁵ Cette redéfinition de la santé et de la maladie a effectivement abouti à une responsabilisation croissante des patient.es au regard de ceux-ci. Loin de pouvoir être encore considéré.es comme passifs/passives, comme les victimes d'une autorité médicale paternaliste et aliénante,³⁶ les patient.es seraient désormais encouragé.es à s'engager activement dans la surveillance prudente de leur santé. Elles/ils seraient ainsi devenu.es des agent.es individuellement investi.es dans la promotion de leur état de santé à travers la consommation d'informations, de services et de produits biotechnologiques disponibles sur le marché globalisé de la santé.³⁷ Cette tendance à la transformation des corps à l'échelle moléculaire

par les technosciences et la déconstruction de la santé en termes de risque, dans laquelle autant le secteur biomédical que les patient.es ont un rôle à jouer, est ultimement une expression du nouveau rapport à la santé caractérisant les sociétés libérales occidentales. Or il nous semble que les pratiques médicales que proposent les banques privées de sang de cordon s'inscrivent précisément dans la continuité de ces transformations. En incitant la mère à imaginer son enfant comme un « malade présymptomatique » (Rose, 2009, p. 9), elles positionnent tout d'abord l'enfant dans un rapport tout à fait spécifique à sa propre santé et à son propre corps.³⁸ Elles dessinent également pour la mère un modèle de participation à la médecine radicalement à rebours des responsabilités et des devoirs qui accompagnaient le « droit à la santé » assuré par l'État-providence, au fondement de l'économie solidaire de tissus biologiques humains. Par la possibilité de consommer ce nouveau service, la nouvelle mère se trouve en effet incitée à « réimaginer qui sont ceux qui se trouvent sous ses propres responsabilités » (Rose, 2007, p. 146). Elle n'est plus encouragée à faire « don » d'un de ses tissus afin de promouvoir la santé collective d'une communauté politique, mais à participer à de nouvelles « pratiques incorporées de mise en culture de soi »³⁹ (Fannin, 2011, p. 348) dans la visée d'assurer le mieux possible (soit le plus prudemment possible) la santé future de son enfant et d'acquérir ainsi son identité de « bonne mère » en puissance.

Au sein de ce secteur, qui s'inscrit selon nous de manière délibérée dans le prolongement de certaines normativités sociales entourant le fait de « bien prendre soin » de son enfant,⁴⁰ la relation entre les parents et leurs enfants semble alors se trouver projetée dans un futur incertain et inconnaissable conceptualisé sous l'angle du risque et qu'il s'agit pour les mères de sécuriser à travers de *nouvelles* formes de consommation. Ces dernières ne prennent en effet pas la forme d'un consumérisme passif et statique, mais se présentent plutôt comme « des formes de travail antérieurement imprévues sur le soi » (McNay, 2009, p. 55). Encouragées à entretenir un rapport très particulier à la santé future de leur enfant d'un côté, et à leurs propres cellules de l'autre, les mères se trouvent ultimement appréhendées par le secteur privé de conservation de sang de cordon comme des collaboratrices engagées de façon proactive⁴¹ dans la coconstruction d'un futur imaginé comme regorgeant d'innovations biomédicales. C'est en étendant la responsabilité des mères au devoir d'offrir à leurs enfants le meilleur futur possible en matière de santé que ce secteur participe à la constitution d'un modèle inédit de participation des individus aux avancées technoscientifiques, dans lequel la frontière entre consommation et production tend à se rétrécir.⁴² Ce modèle se caractérise notamment par un partenariat entre le marché et les mères autour de la maximisation du « capital somatique » de ces dernières.

4.2. De la financiarisation de l'industrie biomédicale à l'émergence d'une nouvelle forme de biocitoyenneté

Ces banques autologues semblent ainsi ouvrir à une « financiarisation » des corps des femmes au sens où ses capacités biologiques régénératives sembleraient pouvoir se multiplier d'elles-mêmes à la mesure de l'avancement futur

des recherches biomédicales en cours et à venir. Cette « mise en valeur » du corps maternel favorise l'émergence de besoins, de responsabilités et de conduites sociales dépassant largement les frontières nettes que les États-providence avaient historiquement posées entre le marché et les tissus vivants, afin de maintenir le corps hors du commerce et d'assurer leurs fonctions de reproduction sociale (Cooper, 2008; Gill et Bekker, 2004).⁴³ Les représentations sociales du corps et de la maternité portées par les banques privées de sang de cordon, ainsi que les pratiques qu'elles favorisent, s'inscrivent en fait particulièrement bien dans la nouvelle financiarisation des sciences du vivant que Rose décrit à travers le terme d'« économie politique de la promesse » (Rose, 2007). Plusieurs chercheuses et chercheurs en sciences sociales ont en effet observé une fusion entre les épistémologies caractérisant le domaine des sciences de la vie et celles propres à l'économie de marché néolibérale, paraissant dès lors se répondre mutuellement (Sunder Rajan, 2006; Cooper, 2006, 2008). En parallèle au tournant néolibéral⁴⁴ des politiques économiques nationales keynésiennes dans les années 1980 et à la mise en place de réformes visant à faciliter la concurrence des entreprises biomédicales, la recherche biomédicale se serait peu à peu mise à fonctionner sur le modèle de l'investissement financier.⁴⁵ Elle se serait trouvée de plus en plus dépendante des perspectives d'innovations médicales et organisée comme un marché dont la « performance » serait « corrélée aux rendements futurs attendus de ses valeurs boursières » (Sunder Rajan, 2006, p. 4). C'est-à-dire que la recherche ne chercherait plus uniquement à produire des « vérités », mais à créer et à mobiliser un capital financier fondé sur des matérialisations futures (*ibid.*, p. 4). De manière concrète, l'« économie de la promesse » décrite par Rose fait également référence à la production de discours médiatiques très largement spéculatifs concernant les possibilités futures qu'ouvriront les recherches en cours. Ces discours seraient notamment produits et diffusés par les nombreux actrices et acteurs cherchant à faire la promotion de projets scientifiques et technologiques dans la visée d'obtenir assez de financements publics et d'investissements privés pour être capables de concourir dans la course à l'innovation et au brevetage (Kaushnik, 2006; Cooper, 2008). L'économie financière (englobant aujourd'hui l'innovation et la recherche) impliquerait donc une autre économie, avant tout « discursive et métaphorique » et à visée performative (Quiet, 2013, p. 273). Cette financiarisation de la recherche biomédicale ouvrirait alors à des relations de collaboration entre différentes actrices et acteurs (États, entreprises, scientifiques et patient.es) autour de l'*espoir* d'extraire de la vitalité du corps une « biovaleur »⁴⁶ (Waldby, 2002), menant à une multiplication des pratiques de capitalisation du vivant (Rose et Novas, 2004). Cet espoir collectivement partagé serait également constitutif de nouvelles normativités sociales autour du « devoir d'agir » individuellement au présent afin de matérialiser concrètement ce futur biologique tant espéré (Rose, 2007, p. 148).

C'est précisément à travers l'émergence de ces nouvelles normativités sociales que l'on verrait se reconfigurer la logique des « projets de citoyenneté » successifs élaborés par les États-nations modernes afin de constituer des citoyen.nes nationaux.⁴⁷ Au courant du XX^e siècle, la citoyenneté avait selon Rose pris la

forme d'une « citoyenneté sociale », fondée sur la détention par les membres de la communauté politique de certains « droits sociaux » tels que les droits au travail, à l'éducation et à la santé.⁴⁸ Or, la manière pour les nations libérales contemporaines de concevoir la citoyenneté semblerait peu à peu s'éloigner de telles considérations. Les États se désengageraient peu à peu de leurs responsabilités à assurer une sécurité minimale à leurs citoyens (en reléguant cette charge aux individus eux-mêmes), tout en investissant de plus en plus « dans l'espoir que certaines caractéristiques génétiques spécifiques de groupes de leurs citoyens puissent potentiellement fournir une ressource précieuse pour la génération de droits de propriété intellectuelle et pour la création d'une biovaleur » (Rose, 2007, p. 133).⁴⁹ Les citoyen.nes seraient de ce fait de plus en plus conduit.es à se concevoir comme les titulaires responsables de ressources biologiques à « valoriser ». Au sein de cette rationalité gouvernementale néolibérale, la citoyenneté ne se confinerait plus à l'ensemble des droits et des devoirs réciproques entre l'État et les individus, mais se définirait par les devoirs moraux que les individus détiendraient au regard de la préservation ou de l'augmentation de leur santé et de celle des membres de leur propre parenté génétique, à travers le ciblage et l'optimisation des capacités vitales du corps. Rose décrit ce phénomène en évoquant l'émergence d'une nouvelle forme de « citoyenneté biologique » (Rose, 2007) correspondant à une mutation du biopolitique dans laquelle le « droit à la santé » assuré par l'État serait reconfiguré en devoir pour les individus de maximiser leur capital santé. Les responsabilités civiques qui accompagnaient le « droit à la santé » instauré par l'État-providence, parmi lesquelles on peut, comme on l'a vu, compter l'acte de redistribuer ses propres *surplus* de vitalité aux autres membres étrangers de la collectivité nationale,⁵⁰ se trouvent dès lors reconfigurées en responsabilités envers soi-même et les personnes auxquelles nous sommes biologiquement lié.es. Dans ce contexte, les citoyen.nes ne sont pas uniquement « laissés à eux-mêmes », en étant conceptualisé.es comme des participant.es proactifs dans l'amélioration de leur propre santé, devant se préoccuper consciencieusement de leur propre devenir biologique individuel en faisant les « choix »⁵¹ les plus judicieux quant à la consommation de produits et de services de santé. Elles/ils doivent également se recollectiviser en fonction de liens et d'affinités somatiques, ouvrant sur de nouvelles formes de « biosocialités »⁵² (Rabinow, 1996) largement médiatisées par le secteur privé. C'est ainsi que les banques autologues de sang de cordon, tout en proposant aux consommatrices d'« investir » activement dans le futur des technosciences, s'inscrivent aussi de façon délibérée dans le prolongement de la solidarité familiale en responsabilisant les mères à l'égard de la santé de leurs enfants. Il nous semble que l'investissement proposé ici aux mères ne doit donc pas être uniquement compris en terme de participation dans l'avenir des technosciences ou la réduction des risques à la santé, mais également comme un investissement identitaire (se rapprocher du statut de la « bonne mère » ou de la « mère idéale »). C'est en jouant sur ces ressorts identitaires et symboliques profonds, en capitalisant sur la solidarité familiale et le sentiment d'obligation morale et affective des mères quant à la protection de leurs enfants, que les banques s'assurent de la « participation » des mères, et font commerce.

CONCLUSION : LA CONSERVATION AUTOLOGUE COMME GARANTE DE LA SOLIDARITÉ FAMILIALE : ENTRE CULPABILISATION MORALE DES MÈRES ET NOUVELLES FORMES D'EXCLUSION (BIO)SOCIALE

Notre analyse comparative des formes de solidarité respectivement promues par les banques publiques et les banques privées de sang de cordon visait à montrer que les banques privées correspondent à une manière spécifique d'imaginer le lien social, radicalement à rebours du « solidarisme »⁵³ (Paugam, 2011) porté par les banques publiques. Elles promeuvent des formes de solidarité familiale où les mères *doivent* – si elles souhaitent préserver le statut identitaire socialement valorisé de la « bonne mère » – agir de la manière la plus prudente possible au regard de la gestion des risques à la santé de leur enfant; cela à travers l'investissement entrepreneurial de leur corporalité dans l'avenir de l'industrie biomédicale. Nous expliquons l'émergence de ces logiques de responsabilité comme la conséquence des réorientations récentes de l'industrie biomédicale et de transformations socioculturelles majeures concernant le rapport des individus à leur santé dans un ordre social néolibéral.

Si les promesses faites aux mères concernant la possibilité d'être un jour en mesure de régénérer la santé de leurs enfants à travers l'utilisation de leurs propres cellules « autologues » sont très loin de se matérialiser dans les années à venir, nous voudrions conclure en considérant la question de la désirabilité de telles thérapies régénératives « personnalisées ». Leur hypothétique concrétisation dans un futur proche et la possibilité pour les mères d'avoir alors directement accès à ces thérapies pour leurs propres enfants nous semblent tout d'abord devoir être débattues au regard des responsabilités immenses que celles-ci viendraient annoncer pour ces mères. En effet, la manière par laquelle les banques autologues jouent sur la relation affective et obligationnelle liant la future mère à son enfant afin de promouvoir leurs services, plaçant la conservation de ces cellules souches dans le prolongement de leur rôle en tant que mères en devenir, ne paraît pas bien éloignée de les blâmer directement pour leur possible refus ou leur inhabilité financière à prendre toutes les précautions possibles afin d'assurer le futur de leurs propres enfants. Dans un contexte socioculturel caractérisé par l'existence de fortes normativités sociales entourant le devoir pour les mères de faire « le mieux possible » afin de prendre soin de leurs enfants, elles-mêmes ancrées dans un idéal de la « bonne mère » comme « prête au sacrifice, domestique et extraordinairement à l'écoute des besoins de ses enfants » (Ladd Taylor and Umansky, 1998, p. 8), l'idée que les mères seraient capables de faire des choix « libres et éclairés », soit « autonomes », concernant la conservation autologue doit être largement questionnée. D'autant plus dans la conjoncture politique actuelle, caractérisée par une *reprivatisation* croissante des risques sociaux dans les foyers et les familles (Bakker et Gill, 2003, Bakker, 2007), se traduisant par une forte responsabilisation des individus dans la prise en charge de leur santé. Selon Bakker et Gill (2003), un des aspects majeurs de la restructuration néolibérale de l'économie politique est effectivement la privatisation croissante de la *reproduction sociale*, c'est-à-dire des institutions et structures

sociales dont toute communauté humaine dépend afin de se maintenir.⁵⁴ Avec la mise en place de programmes d'ajustement structurel et de mesures d'austérité, on assisterait à l'échelle globale à une dissolution des institutions assurant une socialisation des risques et des coûts collectifs liés à la reproduction sociale (par exemple les services publics de santé, de retraite, d'éducation, de transport, de sécurité sociale, d'assurance-emploi, etc.) au profit d'une attente à ce que les individus et les ménages (de concert avec les organisations et les corporations) remplissent ces besoins.⁵⁵ Or, la relégation de cette charge aux ménages – où les femmes se chargent généralement du travail de soins (Bezanson, 2006, p. 4) – viendrait renforcer l'idée que la *reproduction sociale* serait un devoir « naturel » des femmes, s'accomplissant dans la sphère privée. Les femmes devraient donc prendre individuellement en charge les besoins auparavant comblés par les institutions sociales afin d'être considérées comme des citoyennes responsables :

[...] the re-privatisation of social reproduction [which] represents a dual moment: one that returns the work of social reproduction to where it 'naturally' belongs, the household, and, simultaneously, we see women's traditionally caring activities increasingly performed in relationships that are commodified (Bakker, 2007, p. 545).

Au-delà du fait d'inciter les mères à collaborer avec le secteur commercial (et d'exposer ainsi leurs propres vies et celles de leurs enfants à certains risques)⁵⁶ *au nom* de leurs responsabilités maternelles, il nous semble également que ce secteur de conservation « familial » remet en question de manière empirique l'application du principe de justice distributive en santé. Au sens où nous l'entendons, son application devrait minimalement assurer un accès à des soins de santé *de base* (relatifs à la survie) aux personnes n'ayant pas les ressources et les compétences nécessaires afin de rivaliser dans cette course à la « santé parfaite » (Sfez, 1995). Nous nous distancions ici de la position de Rose, pour qui l'émergence de cette nouvelle forme de « citoyenneté biologique » pourrait faire l'objet d'un certain optimisme. Elle serait en effet selon lui susceptible de s'étendre peu à peu jusqu'à mener à une éthique de la responsabilité « sans frontières », liée à une prise de conscience que « tous les êtres humains sur cette planète seraient, après tout, des créatures biologiques » (Rose, 2007, p. 252). L'émergence de cette nouvelle « éthique somatique » (*ibid.*, p. 252) dans le contexte des banques commerciales de conservation de sang de cordon vient cependant, comme nous avons voulu le montrer, illustrer une tout autre réalité. Plutôt qu'à instaurer les conditions nécessaires à une « conscientisation cosmopolite croissante » quant aux « problèmes de santé du pauvre » (*ibid.*, p. 252), elle tend effectivement à produire de nouvelles formes d'exclusion et un renforcement des écarts de santé entre les classes économiquement privilégiées et les autres. Participant de la valorisation de mères responsables et prudentes face à des risques omniprésents à la santé, faisant activement fructifier leur capital biologique au nom du devenir biologique de leur enfant, ces banques les détournent également de l'idée selon laquelle chacun.e serait responsable d'assurer les coûts collectifs de la reproduction de la société et de ses institutions. Concrètement très loin de donner aux familles détenant les ressources financières leur permettant

de contracter leur service les moyens de s'« auto-soigner » de manière autonome,⁵⁷ elles retirent en attendant de manière tout à fait tangible aux malades présents certaines possibilités de guérison, aboutissant à de nouvelles formes d'exclusion (bio)sociale.

Pour conclure, c'est finalement au regard de ces mêmes considérations que la rhétorique plus large d'un nouveau modèle médical « participatif » sur laquelle s'appuie le projet de développer une approche médicale dite « translationnelle » et « personnalisée », nous semble aussi devoir être appréhendée. Ce projet, qui est porté comme on l'a vu par les banques commerciales de sang de cordon, se trouve également partagé par les entreprises privées de génomique personnalisée se multipliant depuis 2004, telles que 23andMe, deCODEme ou encore MyGenome.⁵⁸ Or, l'étiquette de « partenaire » par laquelle les patient.es se trouvent défini.es par ces entreprises nous paraît largement masquer les intérêts commerciaux des institutions biomédicales qui s'en font les représentantes. Les individus auxquels s'adressent ces services « personnalisés » ne sont en effet pas constitués comme de simples consommateurs, mais également comme les coproducteurs de la valeur économique de ces firmes biomédicales.⁵⁹ Ces entreprises dissimulent de ce fait les formes de coercition sociale qui menacent de peser sur les patient.es et leur « autonomie » dans un contexte sociopolitique marqué par une baisse de la prise en charge sociale des problèmes sociaux et une responsabilisation/culpabilisation croissante des individus. C'est en effet *au nom* de leurs responsabilités morales en tant que mères que celles-ci se trouvent intégrées comme des participantes actives au sein « des échelons les plus bas des industries de cellules souches » (Waldby et Cooper, 2013, p. 8). Cette étiquette de partenaire est également susceptible, en faisant miroiter aux individus la possibilité d'obtenir à partir de leurs propres tissus biologiques certains avantages de santé, en matière d'informations ou de nouvelles thérapies, de promouvoir et de généraliser l'idée que ceux-ci seraient de simples « biens », des possessions pouvant être investies dans le développement de futures innovations biomédicales afin de servir un jour les intérêts particuliers de leurs propriétaires. Nous avons voulu montrer dans cet article que cette conception de l'autorité que la/le patient.e est en mesure d'exercer sur ses propres tissus corporels ne constitue pas simplement une entorse à la « dignité inhérente au corps humain » parce qu'elle réfute les principes d'indisponibilité du corps et d'intégrité de la personne. Ces principes visent à tracer une frontière entre les personnes et les choses afin de protéger la personne (comprise comme un tout) contre le risque d'être réduite par autrui – ou encore de se réduire elle-même – au statut de propriété ou d'objet mis au service de buts et d'intérêts subjectifs (Pech, 2001, p. 18). L'affirmation de ces seuls principes semble cependant ne plus suffire. Le développement des biotechnologies (des techniques de transplantation par exemple), en fragmentant le corps en différentes composantes réutilisables, contraint en effet le droit à multiplier les exceptions à ceux-ci (*ibid.*, p. 20). Il reste néanmoins possible de faire appel à une autre interprétation de la dignité humaine, comprise dans sa dimension relationnelle et politique. Envisagée comme un « principe organisateur des libertés » (*ibid.*, p. 27), la dignité humaine peut aussi être définie comme un appel à la pleine responsabilité de la personne dans ses

rapports d'interdépendance aux autres. C'est ainsi au sens où elle ne respecte pas l'exigence de reconnaître autrui en tant qu'autre humanité singulière également vulnérable face aux aléas de la vie, que l'autorité que la consommatrice est en mesure d'exercer sur ses cellules dans le cadre de la conservation privée nous semble aussi constituer une atteinte à la dignité.⁶⁰ De manière concrète, la commercialisation privée des tissus biologiques humains, fondée sur des motivations individualistes (au détriment d'un partage public), semble en effet renvoyer à une tout autre manière de penser et d'organiser politiquement la santé. Celle-ci, de plus en plus assurée par une industrie biomédicale en expansion proposant des services s'adressant aux besoins prétendument spécifiques des individus, nous paraît conduire vers une médecine *stratifiée*, renforçant des inégalités sociales déjà présentes au sein des populations locales. Elle semble exclure les besoins de certaines catégories de la population, privées des modalités d'accès à cette forme soi-disant libre et volontaire de « participation citoyenne » à l'avenir de la médecine, c'est-à-dire, pour parler sans détour, des capacités financières leur permettant de s'offrir ces services.

NOTES

- ¹ Nous faisons référence ici à l'Angleterre, à la France (depuis 1952), à l'Irlande du Nord (depuis novembre 1946), au Luxembourg, à la Slovaquie, à l'ex-Tchécoslovaquie, à la Belgique, au Danemark, à la Finlande et aux Pays-Bas. L'implantation des dons de sang volontaires et gratuits est plus récente à Chypre (1977), en Espagne et à Malte (dans les années 1980) et en Italie, au Portugal, en Grèce et en Estonie (dans les années 1990) (Commission européenne, 2011).
- ² À l'échelle planétaire, 122 pays sur 167 (soit 73% des pays) étaient en 2013 en effet dotés d'une politique nationale de transfusion, incluant 64% des pays à revenu intermédiaire et 41% des pays à revenu faible (Organisation mondiale pour la santé, 2016).
- ³ Cela en analogie, donc, avec les transformations majeures ayant eu lieu en Angleterre au courant du XVI^e siècle, où les terres communautaires et coutumières dont dépendaient les populations rurales ont été redéfinies comme des « champs ouverts » ou des « friches » par le Parlement britannique et transformées en propriétés privées afin de les rendre plus productives (Polanyi, 1944).
- ⁴ Dans leur ouvrage, Clarke et ses collègues (2010) cherchent à mettre au jour les transformations sociologiques qu'ont connues les formes d'organisation, les savoirs et les pratiques des sciences de la vie, avec l'émergence des technologies de l'information à partir du milieu des années 1980. Selon elles, les nouvelles innovations technoscientifiques seraient à l'origine d'une intensification du processus d'expansion de la juridiction médicale sur de nouveaux domaines de la vie, n'opérant non plus seulement à travers un *contrôle* des médecins sur les corps et les vies des patients, mais à travers la *transformation* de ceux-ci à l'échelle moléculaire. Nous développerons davantage autour de certaines de ces transformations plus loin dans l'article.
- ⁵ On pense à des parties d'organes, des embryons, des ovules, des cellules, des lignées cellulaires, etc.
- ⁶ Cette globalisation des échanges est en partie due à la difficulté qu'ont les autorités publiques à réguler à partir de leurs standards traditionnels des « bio-objets » (Webster et coll., 2012) aux frontières échappant souvent aux catégories socioculturelles conventionnelles – telles que les catégories entre la nature et la technique, le vivant et le non-vivant, les applications présentes et les innovations futures, etc.
- ⁷ On pense notamment aux tissus musculaires et osseux, cardio-vasculaires, ophtalmiques, cutanés et reproducteurs, ainsi qu'aux cellules souches de sang de cordon, de moelle osseuse et de sang périphérique.
- ⁸ Ou du moins, au sein de circuits redistributifs visant à les rendre gratuitement disponibles aux autres citoyens, comme c'est le cas aux États-Unis par exemple. Bien qu'ils n'offrent pas d'assurance-maladie publique, les États-Unis sont en effet le pays détenant le plus grand nombre de banques publiques au monde (Dickenson, 2013, p. 90).
- ⁹ Pour une définition de ce terme sociologique, nous renvoyons à la note n°20.
- ¹⁰ Nous avons choisi de mettre le mot *autologue* entre guillemets en raison de l'utilisation abusive qui nous semble être faite de ce terme dans le cas des banques dites autologues de sang de cordon. Le terme autologue signifiant « envers soi-même », soit pour ses propres besoins personnels, les greffes autologues font référence à des greffes réalisées avec un échantillon provenant de son propre corps. Toutefois, nous pensons, à l'instar de l'éthicienne médicale et philosophe Donna Dickenson, que l'échantillon n'appartient pas à l'enfant au moment de sa naissance comme cela est supposé par la plupart des cliniciens et bioéthiciens, et que ce n'est donc pas lui qui le donne, mais bel et bien sa mère, qui l'a « produit » au moment de l'accouchement et a choisi par la suite de le préserver pour son enfant (Dickenson, 2007). Étant donné que c'est la mère qui préserve son échantillon pour son propre enfant, l'utilisation du terme de *don allogénique dirigé* (ou *intrafamilial*) serait ainsi à notre avis plus juste que celle d'*autologue*.

- ¹¹ Notamment les comités consultatifs d'éthique d'Autriche, de Belgique, de Chypre, de France, de Grèce, d'Irlande et d'Italie.
- ¹² Voir par exemple les recommandations publiées par l'American Academy of Pediatrics (2007), The American College of Obstetricians and Gynecologists (2008; 2015), la World Marrow Donor Association (2006), The Royal College of Obstetricians and Gynaecologists (2006), et The Royal College of Midwives au Royaume-Uni (2011). Au Canada, cette coexistence a également généré plusieurs débats, autant parmi des clinicien.nes que des bioéthicien.nes. Un des premiers articles traitant des banques autologues de sang de cordon, publié à la toute fin des années 1990, analyse ainsi les préoccupations des médecins concernant la conservation autologue de sang de cordon, entourant notamment la très faible probabilité que celle-ci soit un jour d'une quelconque utilité (Hass, 1999, cité dans Haw, 2014, p. 9.). C'est d'ailleurs pour cette même raison que The Society of Obstetricians and Gynecologists Canada (SOGC) s'est opposée aux banques privées en 2000, puis de nouveau en 2005. Les bioéthiciens Saginur, Karaboyan et Knoppers ont par la suite publié un article influent examinant les enjeux éthiques et légaux de la conservation de sang de cordon au Canada (Saginur, Karaboyan et Knoppers, 2005). Celui-ci mettait en avant certains enjeux éthiques tels que le flou existant autour du statut juridique octroyé au sang de cordon ou encore l'usage de publicités mensongères par les banques privées de sang de cordon. Il discutait également de la mise en place de régulations permettant de s'assurer que les couples reçoivent des informations exactes concernant la conservation autologue. Depuis cet article, Knoppers et ses collègues se sont toutefois distanciés d'une discussion critique concernant les enjeux éthico-juridiques posés par les banques privées de sang de cordon afin de discuter plutôt de la nécessité de renforcer un programme national de banques publiques de sang de cordon au Canada. Plant et Knoppers (2005) proposent ainsi la mise en place de partenariats public-privé, qui permettraient d'alléger les coûts élevés de la conservation du sang de cordon dans le secteur public et de faire avancer la recherche (Plant et Knoppers 2005, cité dans Haw, 2014, p. 9). Mais, quelle que soit l'analyse que l'on fasse de cette position, il reste que les enjeux éthiques qui avaient été soulevés par Knoppers, Saginur et Karaboyan en 2004, notamment l'usage par les banques privées de pratiques publicitaires douteuses visant à obtenir le consentement éclairé des parents dans des situations où ceux-ci sont particulièrement vulnérables, n'ont pas été discutés de nouveau par la suite (Haw, 2014, p. 9).
- ¹³ Le terme « solidarité » renvoie dans le présent article aux logiques de responsabilité et d'engagement qui sous-tendent différentes formes du « lien » existant entre les personnes (par exemple la solidarité familiale, nationale, etc.), c'est-à-dire aux rapports d'interdépendance qui prévalent entre les personnes au sein de divers espaces socioculturels. Ce terme sera ici utilisé comme un outil heuristique nous permettant de comparer les deux modèles sociaux que valorisent les banques autologues et les banques allogéniques. Il nous permettra de désigner les différentes formes du « lien » auquel nous avons affaire et d'identifier les logiques de réciprocité dans lesquelles elles s'inscrivent.
- ¹⁴ Appréhendés comme des tissus « abandonnés » ayant perdu la capacité à représenter la personne dont ils proviennent, ainsi que toute forme de valeur symbolique ou instrumentale apparente, ceux-ci étaient en effet déjà récupérés par les hôpitaux et les cliniques ayant le droit de les conserver. Ils étaient alors également vendus de manière routinière à des laboratoires de recherche, ainsi qu'à des compagnies pharmaceutiques et cosmétiques en extrayant l'albumine nécessaire à la fabrication de médicaments et de produits de maquillage, sans que le consentement éclairé des mères ne paraisse le moindrement nécessaire (Andrews et Nelkin, 2001, cité dans Waldby et Mitchell, 2001, p. 115).
- ¹⁵ Comme l'expose Dickenson (2007; 2008), son extraction n'est pas réellement réalisée après l'accouchement, mais *durant* celui-ci, souvent avant même que le placenta n'ait été propulsé hors de l'utérus : « Contrary to the cheerful impression given by commercial cord blood banks and echoed by a surprisingly number of otherwise well-informed ethical and legal scholar, the collection of umbilical cord blood takes place during childbirth (in the third stage of labour, after the delivery of the baby and up to the delivery of the placenta). As far as the mother is

concerned, childbirth is not over after the baby is delivered; indeed, the greatest risks to her lie in the third stage, since post-partum haemorrhage is the greatest cause of maternal death [...] » (Dickenson, 2007, p. 87).

- ¹⁶ Voir par exemple cette citation: « They have an advantage over other sources of stem cells [...] because the supply is unlimited, they can be used in autologous or allogeneic situations, they need minimal manipulation, and they raise no ethical concerns » (Gluckman et Vanderson, 2009, p. 454). Du fait que leur extraction ne semble pas poser d'enjeux éthiques particuliers, les cordons et les placentas sont également définis par les chercheuses et chercheurs comme une ressource entièrement disponible et inépuisable de cellules souches : « Cord blood is an unlimited source of hematopoietic stem cells for allogeneic hematopoietic stem cell transplant. Indeed, every minute around 150 babies are born worldwide, rendering umbilical cord blood the most abundant source of human cell material » (Kadereit et Rudolph, 2010, p. 335).
- ¹⁷ Le concept sociologique de « bio-objet », défini par Vermeulen, Tamminen et Webster (2012), fait référence aux nouvelles formes de vie qui sont produites par les biosciences et qui se situent au croisement des catégories anthropologiques du vivant et de l'inerte, du sujet et de l'objet, etc. Le terme « bio-objectification » a selon les auteurs pour fonction heuristique de comprendre comment ces formes de vie sont créées (principalement dans les laboratoires) à travers différentes pratiques sociales et matérielles.
- ¹⁸ Retraçant le parcours des scientifiques impliqués dans le développement de la cryobiologie à partir du début du XX^e siècle, l'historienne H. Landecker montre en effet de quelle manière la première mise en culture cellulaire établie en 1907, née de l'idéal d'ingénierie du vivant chez les chercheuses et chercheurs (notamment d'une détermination partagée à aller « aussi loin que possible » dans l'altération « catastrophique » de la matière vivante sans pour autant que « le fil de la vie ne se brise »), fut le point de départ d'une longue remise en question de notre représentation culturelle du vivant, qui eut pour conséquence de redéfinir le vivant comme un phénomène technologique que l'on peut remodeler à souhait (Landecker, 2007, p. 8).
- ¹⁹ Cela en attribuant des potentialités spécifiques aux cellules souches placentaires, en définissant les modalités de leurs échanges, et en précisant le rapport que les mères sont en mesure d'entretenir avec les produits originaires de leurs propres corps.
- ²⁰ La biopolitique serait, selon Foucault, le résultat des transformations ayant eu lieu dès le début du XIX^e siècle concernant la manière dont l'État exerce son autorité sur les citoyen.nes. La forme d'exercice du pouvoir étatique serait passée d'un pouvoir non plus exercé à travers la mort et le contrôle du territoire (c'est-à-dire la peine de mort ou le maintien des malades contagieux aux portes de la ville et de la communauté), mais à travers la création de la catégorie de *population*, conceptualisant le peuple comme un ensemble d'*êtres vivants*, détenant dès lors des taux de natalité, de mortalité, de fécondité évalués à travers les statistiques et pouvant être *contrôlés*, soit exercée par une rationalisation, un contrôle du corps biologique et de la vie elle-même (Foucault, 1994b, p. 16; 2001, p. 1012).
- ²¹ Caractérisée par « la formulation d'un nouveau droit, d'une nouvelle morale, d'une nouvelle économie, d'une nouvelle politique du corps » (Foucault, 1974, p. 42).
- ²² Voir à titre d'exemple, le nom « *biosurance* » donné à la banque québécoise de la clinique Ovo située à Montréal.
- ²³ Lorsque l'enfant atteindra la majorité, il devra décider s'il veut continuer de préserver l'échantillon qui a été conservé pour lui par sa mère, et ainsi prendre à sa charge les coûts liés à sa conservation.
- ²⁴ La durée maximale de conservation des cellules souches hématopoïétiques est effectivement évaluée à une vingtaine d'années, après quoi celles-ci ne sont plus viables, parce qu'elles perdent leur capacité à se différencier et à reproduire d'elles-mêmes (Broxmeyer et coll., 2011).
- ²⁵ La vaste majorité des usages médicaux du sang de cordon, généralement orientés vers le traitement de la leucémie, sont en effet « allogéniques », notamment en raison du risque que les cellules des patient.es à traiter soient elles-mêmes cancéreuses. Comme l'explique en effet l'American Academy of Pediatrics (2007) : « Cord blood donation should be discouraged

when cord blood stored in a bank is to be directed for later personal or family use, because most conditions that might be helped by cord blood stem cells already exist in the infant's cord blood (ie, premalignant changes in stem cells) » (p. 167).

Les estimations varient entre 5 à 37 sur 100 000 (Comité consultatif de bioéthique de Belgique, 2007). Comme le soulève le Comité consultatif de bioéthique de Belgique (2007), on peut par ailleurs se demander combien de ces 37 personnes pourraient être traitées ultérieurement par une source allogénique (*ibid.*, p. 26). Ces quelques usages de cellules autologues ne sont en fait pas orientés vers le traitement de maladies (contrairement aux nombreuses informations présentées sur les sites Internet des banques autologues), mais afin de traiter les effets des traitements du cancer (Haw, 2014, p. 15). Dans le cas de patient.es atteints de tumeurs cancéreuses solides n'ayant pas leur origine dans le sang ou le système immunitaire (le rétinoblastome ou le neuroblastome par exemple) et ayant été traité.es par radiothérapie, les cellules autologues sont en effet transplantées sur la/le patient.e afin de reconstituer son système immunitaire endommagé (*ibid.*, p. 15). Les chances d'utilisation par un individu du sang du cordon ombilical conservé lors de sa naissance pour le traitement d'un désordre hématoïétique éventuel avant l'âge de 20 ans sont ainsi très minces.

- ²⁶ L'AAP formule par exemple cette recommandation dans son avis : « Physicians should be aware of the unsubstantiated claims of private cord blood banks made to future parents that promise to insure infants or family members against serious illnesses in the future by use of the stem cells contained in cord blood » (AAP, 2007, p. 167). Le ACOG affirme, de son côté : « The routine storage of umbilical cord blood as biological insurance against future disease is not recommended by the American Academy of Pediatrics, given the lack of scientific data to support its use and availability of allogeneic transplantation. Physicians or other professionals who recruit pregnant women and their families for for-profit umbilical cord blood banking should disclose any financial interests or other potential conflicts of interest » (ACOG, 2016, p. 2). La WMDA déclare également : « Today the likelihood that an autologous cord blood unit will be used for transplantation is very low. There is currently no clear proof that these cells will be able to be used for regenerative medicine or to treat other diseases in the future. There are several research studies in the early stages of testing efficacy of autologous cells for some indications. However, at the present time it is impossible to predict the results of research that may affect the future potential use of these cells » (WMDA, 2011, p. 1).
- ²⁷ Notamment le Groupe européen d'éthique (2004), ainsi que les comités consultatifs de l'Autriche (2008), de la Belgique (2007), de Chypre (2004), de la France (2012), de la Grèce (2007), de l'Irlande (2008) et de l'Italie (2007).
- ²⁸ Le comité soutient toujours cette position dans le dernier avis produit en 2012 concernant la conservation autologue du sang de cordon (CCNE, 2012).
- ²⁹ Il semble à ce sujet important de soulever qu'au Canada, on peut déjà compter plus d'une vingtaine de banques privées, dont la banque *OVO Biosurance* et *Cells for Life* au Québec, contre deux banques publiques seulement, *HémaQuébec* et *Alberta Cord Blood Bank*.
- ³⁰ Pouvant être vues, à la fois comme une partie de la personne dont elles sont issues – ou comme des futures personnes, dans le cas des embryons – ou comme des médicaments, selon le potentiel qui leur est attribué, n'étant alors ni vraiment des choses, ni exactement des personnes, ou simultanément, à la fois des personnes et des choses (Pottage, 2004, p. 5).
- ³¹ Rappelons qu'il est légalement interdit dans la plupart des pays de faire commerce des parties détachées de son propre corps, bien qu'un large débat puisse évidemment être ouvert concernant la manière dont la rhétorique du « don altruiste » vient souvent largement masquer, notamment en ce qui concerne le « don altruiste » d'ovocytes, ce qui est en réalité une mise en vente par les individus de leurs produits corporels (Lafontaine, 2014).
- ³² Soit à la mauvaise personne, comme le soulève à juste titre Dickenson, la mère étant la seule impliquée dans les efforts nécessaires à son extraction durant le travail de l'accouchement, et l'identité génétique n'ayant par ailleurs jamais été au fondement d'un quelconque droit de propriété (Dickenson, 2007, p. 94-102).

- ³³ Obligation pour la mère tout d'abord, puis pour le jeune adulte ensuite. À défaut de quoi cette propriété serait transférée systématiquement de l'enfant à la banque, qui pourrait alors l'utiliser comme matériel de recherche ou le vendre. Les banques stipulent aussi leur droit à disposer librement du sang de cordon dans le cas où l'échantillon de sang serait trop petit pour le garder, contaminé ou non exploitable pour une greffe future.
- ³⁴ Ces risques sont notamment liés au manque d'attention que pourrait porter le personnel de santé à la mère et à l'enfant, celui-ci devant simultanément s'occuper de collecter le sang (Dickenson, 2007; 2008). Comme le soutient en effet Dickenson: « Contrary again to the advertising literature, the extraction of cord blood also presents sufficient risks to mother and baby for major professional bodies in obstetrics and gynaecology to have expressed substantial concerns. These qualms can be subdivided into those involving the mother's health and those concerning the best interests of the baby. In addition, both mother and baby may be adversely affected if the attention of delivery room staff is distracted from the primary purpose of a safe delivery » (Dickenson, 2007, p. 87-88).
- ³⁵ De sorte que l'on n'est plus uniquement « sains » ou « malades », mais désormais considérés comme « à bas », « à moyen » ou « à haut risque » d'être malades, soit situés sur un continuum allant de la présence de la maladie avérée à l'objectif de la santé parfaite. Cette transformation doit être située dans la continuité de deux changements épistémologiques ayant eu lieu au courant du XX^e siècle. La redéfinition de ce qu'est la santé par l'Organisation mondiale de la santé (OMS) en 1948, soit non plus comme une absence de maladie, mais comme un « état de complet bien-être fois physique, mental et social » (Préambule à la constitution de l'OMS, 1946), a d'une part opéré un brouillage des frontières entre les notions de bien-être, de confort et de santé. D'autre part, c'est l'émergence, avec la santé publique au lendemain de la Seconde guerre mondiale, d'une nouvelle médecine fondée sur la « surveillance » des « populations normales » (avec le développement de l'épidémiologie et des enquêtes de santé populationnelles) qui opère une reconceptualisation de la santé et de la maladie : auparavant conceptualisées comme des variables discrètes, elles sont redéfinies comme des variables continues). À ce sujet, voir J. Collin (2007).
- ³⁶ Nous faisons ici référence aux critiques que de nombreux sociologues ont fait de la médicalisation et du rapport d'autorité entre médecin et patient.e caractérisant celle-ci, souvent définie comme paternaliste et donc infantilisant, coercitif et aliénant pour les patient.es. Cette représentation d'un.e patient .e passif, non critique vis-à-vis de l'autorité médicale a toutefois été modérée par certain.es auteur.es, analysant la manière dont les patient.es ont aussi parfois participés activement au processus de médicalisation afin de répondre à leurs propres besoins. Voir par exemple l'article de K. Riessman (1989), où l'auteure analyse la participation des femmes à la médicalisation de la naissance ou du symptôme prémenstruel.
- ³⁷ Le développement croissant du tourisme médical nous semble très bien exemplifier les transformations ici décrites. On peut consulter à ce sujet l'article de Whittaker et Leng (2016).
- ³⁸ Les futurs nouveau-nés nous paraissent en effet se trouver placés au cœur d'un projet adultocentriste qui conditionne avant même leur naissance leurs futures attitudes envers leur propre santé.
- ³⁹ Traduction libre de: « Such contracts form distinctive new spatial understandings of the biological body and make possible specific embodied practices of self-cultivation » (Fannin, 2011, p. 339).
- ⁴⁰ L'enfant n'étant pas encore né au moment où la mère contractualise avec une banque commerciale afin d'organiser le prélèvement des cellules de sang de cordon au moment de son accouchement, cette pratique « préventive » s'inscrit dans la continuité des technologies préventives proposées par la médecine prénatale, telles que les tests prénataux. Comme l'a analysé Jeong (2014), ces technologies préventives induisent une responsabilisation des mères et une normalisation de leurs comportements en faisant appel à des représentations sociales largement partagées de la « bonne mère » qui doit faire « tout son possible » afin de protéger au mieux son enfant. Elle montre également de quelle manière ce sont généralement les mères qui sont tenues responsables si l'enfant naît avec une maladie ou un handicap (Jeong, 2014, p. 359).

- ⁴¹ Cela à travers leur contribution financière d'une part, et l'investissement de leur propre corps d'autre part.
- ⁴² Cette « collaboration » repose à la fois sur la consommation par les mères d'un service de conservation autologue, et sur la production par celles-ci d'un « travail clinique » (*clinical labor*) pour reprendre le terme de Cooper et Waldby (2014), étant donné qu'elles fournissent le travail reproductif soutenant cette industrie autologue.
- ⁴³ Pour une définition du terme de reproduction sociale, on peut se référer à la note 52 ci-dessous.
- ⁴⁴ Lorsque nous parlons de « tournant néolibéral », nous faisons référence au processus par lequel les idées portées à partir des années 1930 par Ludwig von Mises, Friedrich von Hayek et d'autres figures de l'École de Vienne, après s'être peu à peu disséminées à l'échelle internationale (d'abord par l'entremise de réseaux intellectuels tels que la Société du Mont-Pèlerin, puis à travers l'essor de l'École de Chicago, notamment autour de la figure de Milton Friedman), ont été *traduites* en politiques économiques nationales concrètes sous l'impulsion de la pression exercée sur les États par les institutions internationales que sont la Banque mondiale, le Fonds monétaire international, l'Organisation mondiale du commerce et l'Organisation de coopération et de développement économique. Il nous semble également que cette traduction politique des idées des premiers intellectuels néolibéraux, en plus d'être une approche politique particulière en matière d'économie, représente également, pour reprendre l'analyse de Foucault, une rationalité politique. Celle-ci transforme activement « la loi du marché », soit la compétition et la concurrence, en principe régulateur de tous les domaines de la vie sociale, se détournant alors de la visée politique de « réparation » du lien social portée par le libéralisme keynésien. Sur le plan des politiques sociales, cela s'exprime par exemple par une responsabilisation croissante des citoyen.nes au regard de leur santé, de leur vieillissement ou de la possibilité de se retrouver sans emploi, soit plus généralement, par une privatisation des *risques*, au détriment de la prise en charge collective de ceux-ci (que visaient à assurer les systèmes de sécurité sociale mis en place dans les années 1930 et 1960).
- ⁴⁵ Le *Bay-Dole Act*, voté par le Congrès américain en 1980 a permis aux universités américaines bénéficiant de fonds fédéraux pour leur recherche, de breveter leurs découvertes et de privatiser ainsi les retombées de la recherche publique à des fins commerciales. Il illustre ainsi particulièrement bien la manière dont les États libéraux se sont peu à peu mis, dans les trois dernières décennies, à collaborer avec le secteur biomédical afin de stimuler leurs propres économies nationales. Cela a rapidement abouti, notamment à travers la décision de la Cour Suprême concernant le procès *Diamond Vs. Chakrabarty* (permettant le brevetage du vivant génétiquement modifié), à un brouillage croissant de la distinction entre les secteurs public et privé (Helmreich, 2008, p. 471). Le vote par le Congrès américain de l'acte de réforme *Private Securities Litigation* en 1995, sous la présidence de Clinton, est un autre exemple concret de la manière dont l'État américain, souhaitant encourager l'innovation biomédicale en stimulant la compétitivité entre les entreprises, a largement contribué à la financiarisation du secteur biomédical. Cet acte a effectivement assuré aux firmes biomédicales et pharmaceutiques une protection légale contre des poursuites judiciaires qui pourraient leur être intentées suite aux faux espoirs engendrés par leurs discours prospectifs (à la condition toutefois que ceux-ci s'accompagnent de mises en garde explicites concernant « la nature incertaine » de ces promesses) (Rose, 2007, p. 90-91).
- ⁴⁶ C. Waldby définit la « biovaleur » (*biovalue*) comme « le surplus de vitalité produit à travers la reformulation biotechnique des processus du vivant. La biotechnologie tente de gagner des actions dans les processus du vivant, de les inciter à augmenter ou de changer leur productivité dans des directions spécifiques, d'intensifier leur capacités à se reproduire d'elles-mêmes et à se conserver » (Waldby, 2002, p. 310).
- ⁴⁷ Selon Rose, que la citoyenneté ait historiquement pu être définie comme la détention de certains droits civils au XVIII^e siècle, comme la capacité à participer au politique à partir du XIX^e siècle, ou encore comme une « citoyenneté sociale » au courant du XX^e siècle, elle aurait continuellement été l'objet d'une construction élaborée par les autorités modernes. Cette construction aurait reposé elle-même sur une *évaluation de la valeur biologique* des indivi-

du informant l'inclusion (et la régulation politique) ou l'exclusion de certains groupes sociaux de la catégorie privilégiée ou protégée de « citoyen » à caractère fortement nationaliste et eugéniste (Rose, 2007, p. 131-132). C'est en ce sens qu'il affirme : « [...] Des présuppositions biologiques, explicitement ou implicitement, ont sous-tendu de nombreux projets de citoyenneté façonnés les conceptions de ce que signifie le fait d'être un citoyen, et fondé les distinctions entre les citoyens actuels, potentiels, problématiques et impossibles » (Rose, 2007, p. 132, traduction libre). Le sociologue fait notamment référence ici aux mécanismes d'inclusion et d'exclusion de la citoyenneté ayant reposé sur une qualification des êtres humains réalisée en fonction de leur sang, de leur ethnicité, de leur physionomie ou de leur sexe, ou encore de l'évaluation de leur intelligence, de leurs capacités morales inhérentes ou de leur bagage génétique.

⁴⁸ Au sujet de cette citoyenneté « sociale », nous renvoyons au chapitre « The Social Citizen » dans Rose (1999).

⁴⁹ On peut, afin de soutenir l'affirmation de Rose concernant la tendance des États libéraux contemporains à concevoir leurs propres populations nationales comme des ressources biologiques précieuses sur le plan économique, mentionner à titre d'exemple la passerelle établie entre la compagnie biopharmaceutique deCODE Genetics et le gouvernement irlandais, ce dernier ayant accepté de « louer » à la firme l'accès aux données génétiques de sa population pendant plus de dix ans dans les années 1990 (Palsson et Rabinow, 1999). On peut également mentionner le financement par le ministère de la Recherche, de la science et de la technologie du Québec de l'organisme Génome Québec (représentant 40 millions de dollars) dans la visée de permettre au Québec de prendre position sur la scène internationale en recherche génomique et de rassurer les investisseuses et investisseurs potentiels à injecter du capital de risque dans les firmes biomédicales québécoises (Bibeau, 2002).

⁵⁰ Si l'État-providence avait en effet entrepris, selon la sociologue M. Cooper, de prendre en charge au milieu du XXe siècle la chronologie de la vie humaine, du début à la fin, la vie des travailleuses et travailleurs et la vie reproductive de la nation, en assurant à chaque citoyen.ne un droit à la vie et à la santé, celui-ci aurait toutefois exigé des citoyen.nes qu'elles/ils donnent en échange à la collectivité leurs vies, leur dévotion à la nation (Cooper, 2008, p. 8).

⁵¹ Le terme « choix » est effectivement à utiliser avec grande précaution puisqu'il est conditionné par le milieu social et économique des personnes, et par les différentes normes et valeurs sociales qui orientent leurs comportements. Comme le remarque également l'anthropologue S. Franklin, les « nouvelles options technologiques produisent un choix forcé » au sens où « une fois que le choix existe, il doit être accepté ou refusé » (Franklin, 1998, p. 108).

⁵² L'anthropologue P. Rabinow (1996) a proposé ce concept afin de décrire les transformations contemporaines des identités sociales autour des savoirs et pratiques des biotechnologies ainsi que les nouvelles collectivités sociales émergentes autour de caractères somatiques ou génétiques partagés.

⁵³ Le « solidarisme » est envisagé par S. Paugam (2011) comme la solidarité sociale propulsée de manière verticale par les institutions de l'État-providence et pensée sur une base nationale. Ce contrat social présuppose donc l'affiliation des individus à une structure sociale et institutionnelle, et non pas simplement des liens sociaux interindividuels.

⁵⁴ Bakker et Gill (2003) s'inscrivent dans un féminisme matérialiste affirmant que c'est le *travail domestique* qui supporte et maintient socialement la production économique. Cela se traduit par le simple fait qu'il est nécessaire que la force de travail soit biologiquement reproduite (par la mise au monde de nouvelles générations de travailleuses et travailleurs, mais également à travers son maintien quotidien à travers la production, l'acquisition, la distribution ou la préparation des moyens de subsistance, incluant notamment la nourriture, le logement, les vêtements et les soins de santé). Bakker et Gill élaborent davantage autour de cette idée en identifiant trois dimensions de la reproduction sociale : 1) la reproduction biologique de l'espèce, impliquant les constructions sociales de la maternité dans différentes sociétés ; 2) la reproduction de la force de travail, impliquant non seulement les moyens de subsistance, mais également l'éducation et la formation ; 3) la reproduction des pratiques sociales liées au soin

(*caring*), à la socialisation et à la satisfaction des besoins humains, qui peuvent être soit privatisées dans les familles, soit socialisées, ou encore en partie privatisées et socialisées (Bakker et Gill, 2003, p. 232).

- ⁵⁵ Les politiques fiscales néolibérales mises en place afin de réduire les déficits de l'État en coupant dans ses dépenses budgétaires ont en effet drastiquement diminué les budgets de la santé et de l'éducation de tous les pays du monde, au nord comme au sud. À ce sujet, voir l'article de G. Sen (1997).
- ⁵⁶ Nous renvoyons à ce sujet aux notes 11 et 28.
- ⁵⁷ Vis-à-vis des technologies médicales et des « surplus » biologiques provenant des corps étrangers de leurs concitoyens.
- ⁵⁸ Projet auquel les banques commerciales de sang de cordon semblent explicitement s'associer comme on l'a vu, mais que les chercheuses et chercheurs en sciences sociales voient plus généralement incarné par les entreprises privées de génomique personnalisée, émergeant depuis 2004, telles que *23andMe*, *deCODEme* ou encore *MyGenome*. Voir par exemple l'article de B. Parry (2013) qui analyse le modèle de participation proposé par l'entreprise de tests génétiques *Knowing MyselfTM*.
- ⁵⁹ Cela en matière d'informations génétiques détenant une valeur potentiellement marchande en ce qui concerne les entreprises de génomique personnalisées, et en matière d'investissement financier (et corporel) dans l'avenir de la médecine régénérative, dans le cas des banques autologues de sang de cordon.
- ⁶⁰ Je tiens à remercier vivement l'un.e des évaluatrices/évaluateurs pour son commentaire avisé concernant les différentes conceptions de la dignité humaine.

BIBLIOGRAPHIE

American Academy of Pediatrics (AAP), « Cord Blood Banking for Potential Future Transplantation », *Pediatrics*, vol. 119, 2007, p. 165-170. En ligne, consulté en jan. 2016 : <http://pediatrics.aappublications.org/cgi/reprint/119/1/165>.

Anderson, Benedict, *Imagined Communities. Reflections on the Origin and Spread of Nationalism*, Londres, Verso Editions/NLB, 1983.

Andrews, Lori et Dorothy Nelkin, *Body Bazaar: The Market for Human Tissue in the Biotechnology Age*, New York, Crown, 2001.

Annas, J. George, « The Waste and Longing – The Legal Status of the Placenta – Blood Banking », *The New England Journal of Medicine*, no. 340, 1999, p. 1521-1524.

Austrian Bioethics Commission at the Federal Chancellery, « Cord blood banking », 19 mai 2008. En ligne, consulté en jan. 2016 : www.bka.gv.at/DocView.axd?CobId=31001.

Bakker, Isabella, « Social Reproduction and the Constitution of a Gendered Political Economy », *New Political Economy*, 2007, vol. 12, no. 4, p. 541-556.

Bakker, Isabella et Stephen (dir.), *Power, Production and Social Reproduction: Human In/Security in the Global Political Economy*, New York, Palgrave MacMillan, 2003.

Bezanson, Kate, *Gender, the State and Social reproduction: Household Insecurity in Neo-liberal Times*, Toronto, University of Toronto Press, 2006.

Bibeau, Gilles, « Géno-business Made in Montréal », *Possibles*, vol. 27, no. 1-2, 2002, p. 127-143.

Boileau, Claire, « Controverses institutionnelles et sociales autour du sang placentaire recueilli à des fins thérapeutiques », *Journal des Anthropologues, Association Française d'Anthropologie*, 2002, p. 123-138.

Boyle, James, « The Second Enclosure Movement and the Construction of the Public Domain », *Law and Contemporary Problems*, vol. 66, no. 1, p. 33-74.

Brown, Nik et Alison Kaft, « Blood Ties: Banking the Stem Cell Promise », *Technology Analysis and Strategic Management*, 2006, vol. 18, no. 3-4, p. 313-327.

Broxmeyer, Hal. E. et coll., « Hematopoietic Stem/Progenitor Cells, Generation of Induced Pluripotent Stem Cells, and Isolation of Endothelial Progenitors from 21 to 23.5 Years Cryopreserved Cord Blood », *Blood*, 2011, vol. 117, p. 4473-4777.

Canguilhem, Georges, *Le Normal et le Pathologique*, Paris, PUF, « Collection Galien », 1979 [1966].

Castel, Robert, « De l'intégration sociale à l'éclatement du social : l'émergence, l'apogée et le départ à la retraite du contrôle social », *Revue Internationale d'action communautaire*, vol. 20, no. 60, 1988, p. 67-78.

Clarke, Adèle, Mamo, Laura, Ruth Fosket, Jennifer, Fishman, Jennifer et Janet K. Shim, *Biomedicalization. Technoscientific Transformations of Health, Illness, and US. Biomedicine*, Durham, Duke University Press, 2010.

Collin, Johanne, « Relations de sens, relations de fonction : risque et médicament », *Sociologie et sociétés*, vol. 39, no. 1, 2007, p. 99-122.

Commission européenne, « Deuxième rapport sur le don volontaire et non rémunéré de sang et de composants sanguins », *Rapport de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des Régions*, Bruxelles, mars 2011. En ligne, consulté en jan. 2016 : http://ec.europa.eu/health/blood_tissues_organs/docs/blood_report-donation_fr.pdf.

Comitato Nazionale per la Bioetica, « Mozione del Comitato Nazionale per la Bioetica sulla raccolta e la conservazione di cellule staminali derivate da cordon ombelicale », 13 juill. 2007. En ligne, consulté en jan. 2016 : www.governo.it/bioetica/mozioni/mozionecordonali.pdf.

Comité consultatif d'éthique, science de la vie et de la santé (CCNE), « Sur les biobanques de sang de cordon ombilical en vue d'une utilisation autologue ou en recherche », Opinion, Avis no. 74, 2002. En ligne, consulté en jan. 2016 : <http://www.ccne-ethique.fr/fr/publications/les-banques-de-sang-de-cordon-ombilical-en-vue-dune-utilisation-autologue-ou-en#.VVNQIF6yU7s>.

—, « Utilisation des cellules souches issues du cordon ombilical, du cordon lui-même et du placenta et leur conservation en biobanques. Questionnement éthique », Avis no. 117, 2012. En ligne, consulté en jan. 2016 : <http://www.ccne-ethique.fr/fr/publications/utilisation-des-cellules-souches-issues-du-sang-de-cordon-ombilical#.VVNQWV6yU7s>.

Comité consultatif de bioéthique de Belgique, « Avis n° 42 du 16 avril 2007 relatif aux banques de sang de cordon ombilical ». En ligne, consulté en jan. 2016 : <http://www.health.belgium.be/fr/avis-ndeg-42-banques-de-sang-de-cordon-ombilical>.

Commission Européenne, « Deuxième rapport sur le don volontaire et non rémunéré de sang et de composants sanguins », *Rapport de la Commission au Parlement Européen, au Conseil, au Comité économique et social Européen et au Comité des Régions*, Bruxelles, mars 2011. En ligne, Consulté en jan. 2016: http://ec.europa.eu/health/blood_tissues_organs/docs/blood_report-donation_fr.pdf.

Cooper, Melinda et Waldby, Catherine, *Clinical Labor: Tissue Donors and Research Subject in the Global Bioeconomy*, Durham, N. C., Duke University Press, 2014,

Cooper, Melinda, *Life as Surplus: Biotechnology & Capitalism in the Neoliberal Era*, Seattle et Londres, University of Washington Press, 2008.

—, « Resuscitations: Stem Cells and the Crisis of Old Age », *Body and Society*, vol. 12, no. 1, 2006, p. 1-23.

Cyprus National Bioethics Committee, « Opinion of the Cyprus National Bioethics Committee Regarding Bioethical Dilemmas of Umbilical Cord Blood Banking by Private Profit Making Companies », 27 sept. 2004. En ligne, consulté en jan. 2016 : www.bioethics.gov.cy/law/cnbc/cnbc.nsf/all/b5d91dc497a2eb49c2257308004b46b2?opendocument.

Dickenson, Donna, *Me Medicine vs. We Medicine. Reclaiming Biotechnology for the Commons Good*, Columbia University Press, 2013.

- , *Body Shopping. The Economy Fuelled by Flesh and Blood*, Oneworld, Oxford, 2009.
- , *Property in the Body : Feminist Perspectives*, Cambridge, Law, Medicine and Ethics, Cambridge University Press, 2007.
- Egan, Mary Ellen, « Cell Makers : Biotechs ViaCell and Aastrom Harvest and Grow Stem Cells », *Forbes*, 166, 2000, p. 220-222. En ligne, consulté en jan. 2015 : <http://www.forbes.com/forbes/2000/1211/6615220a.html>.
- European Technology Platform on Nanomedicine, *Vision Paper and Basis for a Strategic Research Agenda for Nanomedicine*, Luxembourg, Office for Official Publications of the European Communities, 2005.
- Fannin, Maria, « Personal Stem Cell Banking and the Problem with Property », *Social & Cultural Geography*, vol. 12, 2011, p. 339-356.
- Foucault, Michel, « Les mailles du pouvoir » (1981), in *Dits et écrits*, tome IV, no. 297, Paris, Gallimard, 1995.
- , « La fonction politique de l'intellectuel » (1976), in *Dits et écrits*, tome III, Gallimard, Paris, 1994a.
- , « Crise de la médecine ou crise de l'anti-médecine? » (1976), in *Dits et Écrits*, tome III, Gallimard, Paris, 1994b.
- , *The History of Sexuality. Vol. I: An Introduction*, trad. R. Hurley, New York, Vintage, 1980.
- Fontaine, Philippe, « Blood, Politics and Social Science: Richard Titmuss and the Institute of Economic Affairs, 1957-1973 », *Isis*, no. 93, 2002, p. 401-434.
- Franklin, Sarah et Margaret Lock, (dir.), « Animation and Cessation : The Remaking of Life and Death », in *Remaking Life and Death. Towards an Anthropology of the Biosciences*, Santa Fe, NM, School of American Research Press, 2003, p. 3-22.
- Franklin, Sarah, « Making Miracles: Scientific Progress and the Facts of Life », in Franklin, Sarah et Hélène Ragonne, (dir.), *Reproducing Reproduction*, Philadelphia, University of Pennsylvania Press, 1998, p. 102-117.
- Gabe, Jonathan, « Medicalization », in Gabe, J. et L. F. Monaghan, (dir.), *Key Concepts in Medical Sociology*, 2^e éd., London, Sage, 2013, p. 49-53.
- Génome Québec, consulté en jan. 2015 : www.genomequebec.com.
- Gill, Stephen et Isabella Bekker, *Power, Production and Social Reproduction. In/security in the Global Political Economy*, New York, NY, Palgrave MacMillan, 2004.
- Gluckman, Elaine et Rocha, Vanderson, « Cord Blood transplantation: state of the art », *The Hematology Journal*, vol. 94, no. 4, 2009, p. 451-454.
- Godbout, Jacques et Caillé, Alain, *L'esprit du don*, Paris, La Découverte, 1992

Groupe européen d'éthique des sciences et des nouvelles technologies (GEE), « Les aspects éthiques des banques de sang de cordon ombilical », *Avis no. 19*, 2004.

Haw, Jennie, « From the Incinerator to the Bank: A Feminist Qualitative Study of Private Cord Blood Banking in Canada », thèse de doctorat, Département de sociologie, York University, 2014.

Helmreich, Stefan, « Species of Biocapital », *Science as Culture*, vol. 17, no. 4, 2008, p. 463-478.

Hoeyer, Klaus, « Anthropologie des objets-frontières », *Sociologie et Société*, vol. 42, no. 2, 2010, p. 67-89.

Holden, C., « Banking a Cord Earn Interest », *Science*, no. 262, 1993, p. 1511.

Irish Council for Bioethics, « Stem Cell Research: Hope or Hype? Exploration of Ethical Questions », 2008. En ligne, consulté en jan. 2016 : www.bioethics.ie/uploads/docs/stemcell.pdf.

Jeong, Yeonbo, « Scientific Motherhood, Responsibility and Hope : Umbilical Cord Blood Banking in South Korea », *New Genetics and Society*, vol. 33, no. 4, 2014, p. 349-369.

Kadereit, Suzanne et Rudolph Gerarald, *Cord Blood Stem Cells, A Future for Regenerative Medicine?*, Singapore, World Scientific, 2010.

Kohler, Riessman C., « Women and Medicalization: A New Perspective », *Social Policy*, vol. 14, 1983, p. 3-18.

Kline, R., « Whose Blood is it Anyway? », *Scientific American*, no. 284, 2001, p. 42-49.

Lafontaine, Céline, *Le corps marché*, Paris, Seuil, « La Couleur des Idées », 2014.

Lafontaine, Céline et Mathieu Noury, « De la nano-médecine à la nano-santé : vers un nouveau paradigme biomédical », *Socio-anthropologie*, n° 29, 2014, p. 13-38.

Landecker, Hannah, *Culturing Life: How Cells Became Technologies*, Massachusetts et Londres, Harvard University Press, Cambridge, 2007.

Le Breton, *L'adieu au corps*, Paris, Métailié, 1999.

Mauss, Marcel, *Essai sur le don. Forme et raison de l'échange dans les sociétés archaïques*, PUF, Collection Quadrige, 2012 [1924].

McNay, Lois, « Self as Enterprise: Dilemmas of Control and Resistance in Foucault's 'The Birth of Biopolitics' », *Theory, Culture and Society*, vol. 26, no. 6, 2009, p. 55-77.

Moulin, Anne-Marie, « Le corps face à la médecine », in Corbin, Alain, Courtine, Jean-Jacques et Georges Vigarello, (dir.), *Histoire du corps. Les mutations du regard, le XX^e siècle*, Paris, Seuil, « L'univers historique », 2005, p. 15-70.

National Bioethics Commission – Greece, « Opinion on Umbilical Cord Blood Banking », 1^{er} avril 2007. En ligne, consulté en jan. 2016 : www.bioethics.gr/media/pdf/recommendations/opinion_uc_eng.pdf.

Novas, Carlos, « The Political Economy of Hope : Patient's Organizations, Science and Biovalue », *Biosocieties*, vol. 1, no. 3, 2006, p. 289-305.

Organisation mondiale pour la santé, « Sécurité transfusionnelle et approvisionnement en sang », *aide-mémoire n° 279*, juillet 2016. En ligne, consulté en sept. 2016 : <http://www.who.int/media-centre/factsheets/fs279/fr/>.

Organisation pour la coopération et le développement économique (OCDE), *The Bioeconomy to 2030. Designing a Policy agenda*, 2009. En ligne, consulté en nov. 2015 : <http://www.oecd.org/futures/long-termtechnologicalsocietalchallenges/thebioeconomyto2030designingapolicyagenda.htm>.

Organisation mondiale pour la santé, « Sécurité transfusionnelle et approvisionnement en sang », *aide-mémoire n° 279*, juillet 2016. En ligne, consulté en sept. 2016 : <http://www.who.int/media-centre/factsheets/fs279/fr/>

OVO Biosurance, consulté en jan. 2015 : <http://www.cliniqueovo.com/en/ovo-biosurance/>.

Palvia, Marie-Luce, « La découverte de la dignité humaine », in M-L. Palvia et T. Revet, (dir.), *La dignité de la personne humaine*, Paris, Economica, 1999, p. 3-23.

Parry, Bronwyn, « Knowing Myself: Personalized Medicine and the Economization of Prospective Knowledge about Bodily Fate », in Meusburger, Peter, Glüker, Johannes et Martina el Meskioui, (dir.), *Knowledge and the Economy*, New York et Londres, Springer, 2013, p. 157-171.

Paugam, Serge, « Conclusion : Vers un nouveau contrat social? », in *Repenser la Solidarité*, Paris, PUF, 2011, p. 949-980. En ligne, consulté en jan. 2015 : <http://www.cairn.info/repenser-la-solidarite—9782130588757-page-949.htm>.

Palsson, Gisli et Paul Rabinow, « Iceland: The Case of a National Human Genome Project », *Anthropology Today*, vol. 15, no. 5, 1999, p. 14-18.

Pech, Thierry, « La dignité humaine. Du droit à l'éthique de la relation », *Éthique publique*, vol. 3, n° 2, 2001. En ligne, consulté en nov. 2016 : <http://ethiquepublique.revues.org/2526>

Plant, Margo et Bartha Maria Knoppers, « Umbilical Cord Blood Banking in Canada: Socio-Ethical and Legal Issues », *Health Law Journal*, vol. 13, 2005, p. 187-212.

Polanyi, Karl, *La grande transformation, aux origines politiques et économiques de notre temps*, Paris, Gallimard, 1983 [1944].

Pottage, Alain, « Introduction: The Fabrication of Persons and Things », in Pottage, Alain et Mundy, Martha (dir.), *Law, Anthropology, and the Constitution of the Social: Making Persons and Things*, Cambridge, Cambridge University Press, 2004, p. 1-39.

Préambule à la constitution de l'Organisation mondiale de la santé, conférence internationale sur la santé, New York, 19-22 juin 1946.

Rabinow, Paul, *French DNA: Trouble in Purgatory*, Chicago, University of Chicago Press, 1999.

—, « Artificiality and Enlightenment: From Sociobiology to Biosociality », in *Essays on the Anthropology of Reason*, Princeton University Press, 1996.

Rose, Nikolas, *The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-First Century*, Princeton, NJ, Princeton University Press, 2007.

—, « The Social Citizen », in *Powers of Freedom: Reframing Political Thought*, Cambridge, Cambridge University Press, 2004, p. 133-136.

Rose, Nikolas et Carlos Novas, « Biological Citizenship », in Ong, Aihwa et Stephen Collier, (dir.), *Global Assemblages, Technology, Politics and Ethics as Anthropology Problems*, Oxford, Blackwell Publishing Ltd, 2004, p. 439-464.

Rose, Nikolas et Paul Rabinow, « Biopower Today », *BioSocieties*, vol. 1, no. 2, 2006, p.195-218.

Saginur, Madelaine, Kharaboyan, Linda et Bartha Maria Knoppers, « Umbilical Cord Blood Stem Cells: Issues with Private and Public Banks », *Health Law Journal*, vol. 12, 2004, p.17-34.

Santoro, Pablo, « From (Public?) Waste to (Private?) Value. The Regulation of Private Cord Blood Banking in Spain », *Science & Technology Studies*, vol. 22, no. 1, 2009, p. 16-12.

Sen, Gita, « Globalization, Justice and Equity: A Gender Perspective », *Development*, vol. 40, no. 2, 1997, p. 21-26.

Sfez, Lucien, *La santé parfaite : critique d'une nouvelle utopie*, Paris, Seuil, 1995.

Society of Obstetricians and Gynaecologists Canada, « Umbilical Cord Blood Banking: Implications for Perinatal Care Providers », *SOGC Clinical Practice Guidelines*, vol. 156, mars 2005. En ligne, consulté en ligne en sept. 2016 : <http://sogc.org/wp-content/uploads/2013/01/gui156ECPG0503wDisclaimer2.pdf>.

Steiner, P., « Don de sang et don d'organes : le marché et les 'marchandises fictives' », *Revue française de sociologie*, vol. 42, no. 2, 2001, p. 357-374.

Sunder Rajan, Kaushnik, *Biocapital: The Constitution of Post-Genomic Life*, Durham, NC, Duke University Press, 2006.

The American College of Obstreticians and Gynecologists, « Umbilical Cord Blood Banking », *Committee Opinion*, no. 648, déc. 2015. En ligne, consulté en jan. 2016 : <https://www.acog.org/-/media/Committee-Opinions/Committee-on-Genetics/co648.pdf>

—, « Umbilical Cord Blood Banking », *Committee Opinion*, no. 399, fév. 2008.

Titmuss, Richard, *The Gift Relationship: From Human Blood to Social Policy*, London, George Allen and Unwin, 1970.

Tutton, Richard et Barbara Prainsack, « Enterprising or Altruistic Selves? Making up Research Subjects in Genetics Research », *Sociology of Health and Illness*, vol. 33, no. 7, 2011, p. 1081-1095.

Vermeulen, Niki, Tamminen, Saka et Andrew Webster, *Bio-objects: Life in the 21st Century*, Ashgate, Serie Theory, Technology and Society, 2012.

Waldby, Catherine et Melinda Cooper, « From Reproductive Work to Regenerative Labour: The Female Body and the Stem Cell Industries », *Feminist Theory*, vol. 11, no. 3, 2013, p. 3-22.

Waldby, Catherine et Robert Mitchell, *Tissue Economies. Blood, Organs and Cell Lines in Late Capitalism*, Durham et Londres, Duke University Press, 2006.

Waldby, Catherine, « Umbilical Cord: From Social Gift to Venture Capital », *Biosocieties*, vol.1, no.1, 2006, p. 55-70.

—, « Stem Cells, Tissue Cultures and the Production of Biovalue », *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine*, vol. 6, no. 3, 2002, p. 305-323.

Whittaker, A. et H. Leng, « 'Flexible Bio-citizenship' and International Medical Travel: Transnational Mobilities for Care in Asia », *International Sociology*, vol. 31, no. 3, 2016, p. 1-19.

World Marrow Donor Association (WMDA), « Annual Report for Cord Blood Banks 2013 ».

—, « Policy Statement for the Utility of Autologous or Family Cord Blood Unit Storage », mai 2006. En ligne, consulté en jan. 2016 : <http://bloodcell.transplant.hrsa.gov/cord/files/wmdapolicystatement.pdf>.

Zola, Irving, K. « Medicine as a Institution of Social Control », *Sociology Review*, 1972, vol. 20, no.4, 1972, p. 487-504.

DOSSIER**METAETHICS, NORMATIVITY, AND VALUE****HICHEM NAAR**

VISITING ASSISTANT PROFESSOR, DEPARTMENT OF PHILOSOPHY, UNIVERSITY OF NEBRASKA AT OMAHA

MICHELE PALMIRA

POSTDOCTORAL FELLOW, DEPARTMENT OF PHILOSOPHY, UNIVERSITY OF BARCELONA

INTRODUCTION

Some of the central questions of moral philosophy revolve around the nature, representation, and apprehension of moral concepts and properties, such as *good*, *permissible*, *wrong*, and the relation they bear to one another, if any. The list of questions taken up in contemporary analytic moral philosophy is familiar enough: “Are moral values and norms objective or subjective?” “Is there any link between what’s good/bad and what’s obligatory/forbidden, and if so, which of the two pairs should come first?” “Are moral values and normative facts reducible to nonmoral/nonnormative facts?” “Are moral judgements expressive of cognitive states, or do they rather express our affective non-cognitive responses to morally relevant actions?”

Answering these questions amounts to accomplishing the difficult task of unifying metaphysical, semantic, normative, and psychological considerations. The articles contained in this special issue jointly rise to this unity challenge by approaching the domains of the evaluative and the deontic from several interconnected perspectives.

In “Goodness: Attributive and Predicative,” Michael-John Turp investigates the celebrated distinction between *attributive* and *predicative* uses of “good” made by Peter Geach and subsequently deployed by authors such as Philippa Foot and Judith Jarvis Thomson. According to Geach, our uses of “good” are chiefly attributive—that is, they require specification of the kind of thing that is said to be (or not to be) good, as in “This is a good move,” “*The Broom of the System* is a good book,” and so on. Geach famously maintained that the absence of a predicative reading of “good” has far-reaching metaethical implications: we should stop asking the classical metaphysical question philosophers have been after since G. E. Moore—i.e., “What is good?”—and instead ask: “What is it to be a good human being?” Turp contends that acknowledging the linguistic datum

that several uses of “good” are attributive does not *ipso facto* make the traditional question of what good *simpliciter* is meaningless or philosophically insignificant. His argumentative strategy unfolds in three steps. First, by drawing an analogy with the attributive evaluation carried out in the assessment of artefacts, Turp stresses that an adequate account of attributive goodness be functional in kind, just like the goodness of a spade is determined on the basis of its function, which goes over and above the designer’s intentions and the present and personal concerns of the user. Secondly, Turp evaluates the prospects of a neo-Aristotelian functional account of natural goodness, notably that by Philippa Foot. This account maintains that, in order to establish what’s good for us, we should seek to discover what traits are conducive to evolutionary fitness, or what way of life serves our interests best. While Turp does not directly take issue with such an account of attributive goodness, he emphasizes an important asymmetry between human beings and other living beings: we have a capacity for rational reflection that enables us to pursue an assessment of the value of whatever conception of attributive goodness is recommended by a naturalistically oriented approach. Given our capacity for rational reflection, the question whether we reflectively endorse a life in which we assign priority to our self-interest will remain nonetheless open for us. That is to say, we can meaningfully ask: “Is a life that is good in the attributive sense a life that is worthy of pursuit?” (p. 81). This, Turp maintains, amounts to asking a question about the predicative goodness of a certain conception of the (attributively) good life. In this way, Turp vindicates the philosophical legitimacy of inquiring into the question of what the moral good *simpliciter* is.

Toni Rønnow-Rasmussen’s starting point in “On-Conditionalism: On the Verge of a New Metaethical Theory” is the fact that some people, metaethicists and laypersons alike, justifiably believe that they don’t know whether value claims are expressive of cognitive states, such as belief, or conative states, such as preference. This meta-metaethical preoccupation leads Rønnow-Rasmussen to lay the foundations for a new metaethical theory, what he calls *On-Conditionalism*, for those speakers (and thinkers) who have no settled yes-or-no opinion on the longstanding debate between cognitivism and noncognitivism. On-Conditionalism maintains that when these speakers make a value claim such as “Pleasure is good,” such a claim is expressive of the cognitive state *that pleasure is good on condition that goodness exists*. Importantly, Rønnow-Rasmussen explores the question of the nature of this conditional belief. He expresses scepticism towards reducing such an attitude as a belief to a conditional in the form “you ought/ought not to believe that x is good, if you learn that goodness exists/does not exist” (p. 97). Rønnow-Rasmussen notices that “I can go on believing that x is good on condition that goodness exists even if I believe that goodness does not exist” (p. 97), so it would be a mistake to say that learning that goodness doesn’t exist forbids me from believing that something is good *conditional* on the existence of goodness. This leads Rønnow-Rasmussen to explore two possible nonreductionist accounts. The first draws on the well-tried distinction between occurrent and dispositional states, and claims that a conditional belief is a dispositional state with a multiple propositional content. While

expressing sympathy towards this proposal, Rønnow-Rasmussen points out that it needs to be supplemented with an account of the occurrent states that speakers are in while making value claims and being agnostic about the cognitivism vs. noncognitivism debate. He therefore puts forward the hypothesis that value claims (1) express beliefs with a categorical content that are the product of a conditional dispositional state, (2) play an explanatory role in precautionary contexts, and (3) are not causally dependent on those beliefs on which nonconditional categorical evaluative beliefs are taken to depend (e.g., the belief that goodness exists). Rønnow-Rasmussen claims that conceiving of the central cognitive attitude introduced by On-Conditionalism in these terms captures the occurrent sense of conditional belief and has the potential to provide a new version of internalism about motivation to the effect that our moral claims are conditionally motivating up to the addition of some external factor.

In “Two Conceptions of Practical Reasons,” Christoph Hanisch contrasts Joseph Raz’s chiefly realist account of practical reasons with Christine Korsgaard’s constructivist account—in relation to various issues in practical philosophy and action theory. While Raz takes reasons for action to be facts that obtain independently of our agency, Korsgaard takes them to be facts generated by constitutive features of the deliberative standpoint. Principles, thereby, play different roles in deliberation on the two accounts. On Raz’s view, the role of principles is merely “to guide the activity of organizing and systematizing ‘normative facts’ [e.g., reasons] and ‘aspects of the world’ in the correct manner” (p. 113). By contrast, Korsgaard takes principles to be prior to reasons in virtue of the fact that our identities as rational agents are themselves constituted by principles, guiding, as they do, our practical reflection. After introducing these two accounts of practical reasons, Hanisch suggests that, notwithstanding appearances to the contrary, the accounts can in fact be combined in an interesting way. Drawing an analogy between practical reasons and placebo drugs, he argues that, even if practical reasons may ultimately derive from our agency, they must be seen as independent of us in order to “exhibit first-personally perceived authority and normative force” (p. 115). A certain illusion—the illusion of the independence of practical reasons—may be necessary for practical reasons to play their role in the constitution of our agency as construed by constructivists. The assumptions we make as deliberative agents, in other words, may be—and perhaps should be—different from the claims we accept when we do metaethics: “As practical deliberators and agents we need Razian realism about the reasons we have, even if, on a different level of reflection, we might have to acknowledge them as originating in our subjective self-understanding” (p. 115). Hanisch ends his discussion with the issue of the relationship between practical reasons and motivation. In particular, he puts forward a view of how agents are guided by principles that does not fall prey to the worry that the constructivist account of rational motivation is too cognitively demanding. In addition, inspired by Hegel, Hanisch argues for a view of action based on reasons, which allows for practical deliberation to occur *after* the action was performed, in turn departing from the classical view that reasons are the causes of actions of this sort.

In “A Stringent but Critical Actualist Subjectivism about Well-Being,” Stéphane Lemaire tackles the issue of what version of subjectivism about well-being someone attracted to this view should accept. According to subjectivism about well-being, “something is good for an individual if and only if this individual has a certain type of pro-attitude A toward this object under conditions C” (p. 134). One central question is then how best to characterize the relevant conditions. In particular, should the relevant pro-attitudes be ones the subject actually possesses or ones the subject would possess in some counterfactual, idealized conditions? In his article, Lemaire defends a version of *actualism*, according to which it is actual pro-attitudes that matter in determining what is good for one. To this end, Lemaire lays out three desiderata for a theory of well-being. First, an adequate theory should accommodate the fact that what is good for one often departs from what one happens to desire (to believe to be good, etc.), a desideratum that is typically taken to motivate both nonactualist and objectivist views of well-being. Second, an adequate theory should not be such as to alienate people from what is good for them. For something to be good for one, it should somehow matter to one. In contrast to the previous desideratum, this desideratum is usually taken to motivate subjectivism in general and actualist subjectivism in particular. To these desiderata—both of which are commonly accepted in the subjectivist literature—Lemaire adds a third one: “At least in some cases, it seems unacceptably paternalistic to act toward someone as if something that matters to him or her counts for nothing” (p. 135). According to Lemaire, we need to accept the claim that what matters to someone counts towards his or her well-being in order to explain why certain paternalistic interventions are wrong—why, for instance, it is wrong to insist that a certain career would not be good for one in cases in which one is very enthusiastic about it and deeply cares about pursuing it. Lemaire goes on to argue that extant actualist versions of subjectivism fail to satisfy the two antipaternalistic desiderata, and that a more “stringent” form of actualism—one that does not resort to any idealization strategy—is called for. According to Lemaire, an adequate actualism about well-being should not discard “badly informed” pro-attitudes as irrelevant, let alone detrimental, to what is good for one. The more stringent the actualism, however, the less apt it appears to deal with the first—normative—desideratum. Lemaire ends his article by outlining a version of actualism—Stringent Critical Actualism—which, while focusing entirely on actual pro-attitudes, makes room for a normative assessment of so-called ‘defective’ attitudes.

In “Les attitudes appropriées *verbatim*” (in French), Julien Deonna and Fabrice Teroni discuss a popular way of understanding evaluative concepts—namely, that to think of an object as having a certain value is to think of it as meriting a certain kind of response. On this sort of view, therefore, evaluative concepts such as GOOD, BAD, ADMIRABLE, and so on are analyzed in terms of appropriate attitudes. To be successful, Deonna and Teroni suggest, such an analysis should meet two challenges. On the one hand, for it to genuinely reductive, a Fitting-Attitude (FA) analysis should not appeal to attitudes whose possession requires a mastery of the evaluative concepts to be analyzed. It should not be the case, in other words, that to think of something as meriting a certain kind of

response presupposes thinking of it as being good, bad, admirable, and so on. This is what Deonna and Teroni call the “psychological challenge.” On the other hand, the account must meet a normative challenge: it should not understand the normative notion of appropriateness or merit itself in terms of evaluative notions. Deonna and Teroni’s aim is to provide a version of the FA analysis that meets both of these challenges. After arguing that an adequate FA analysis should appeal not only to generic attitudes—such as love and hate, desire and aversion, etc.—but also to more specific ones such as emotions (for this would allow the advocate of the FA analysis to elucidate *thick* evaluative concepts), they go on to discuss contemporary approaches to emotions that may be combined with an FA analysis. According to the dominant form of such approaches, the relation between emotions and evaluative properties is that the former somehow put us in touch with the latter. Although there are various ways to spell out this idea, a natural way to interpret it is to say that an emotion *represents* a certain kind of evaluative property. The first task is then to develop a version of this view that meets the psychological challenge. This is where the perceptual theory of emotions comes in, as it claims that, in order to experience an emotion with a certain evaluative content (e.g., “This is disgusting”), the subject need not possess the evaluative concept one would deploy to specify that content. Although this may appear to be a step in the right direction, Deonna and Teroni note that what we should be interested in is not so much emotions themselves, but our *concepts* of them. So even if experiencing an emotion need not require the possession of the relevant evaluative concepts, having an idea of that emotion may. And this claim, Deonna and Teroni go on to argue, is something the perceptualist cannot avoid easily. Following recent works, Deonna and Teroni argue that emotions do not possess an evaluative content at all—hence do not represent values—but constitute distinctively evaluative stances or attitudes that can be felt towards the very same content. Thus, one can be afraid, glad, or sad regarding the same state of affairs. Deonna and Teroni argue that this “attitudinal theory” can be combined with an FA analysis in a way that meets the psychological and the normative challenges.

The last contribution of the present special issue is the first full French translation of Peter Railton’s ground-breaking article “Moral Realism” (1986). As is well known, Railton defends a naturalist reductionist version of realism about moral properties. Railton’s article contains a discussion of all the main questions analytic metaethicists have been wrestling with for more than one hundred years. So, we suggest using it as a reference point to navigate the five new contributions of this special issue and gauge their individual merits in advancing an understanding of the good and the right.

GOODNESS: ATTRIBUTIVE AND PREDICATIVE

MICHAEL-JOHN TURP

LECTURER, DEPARTMENT OF PHILOSOPHY, UNIVERSITY OF CANTERBURY

ABSTRACT:

There is little consensus concerning the truth or reference conditions for evaluative terms such as “good” and “bad.” In his paper “Good and Evil,” Geach (1956) proposed that we distinguish between attributive and predicative uses of “good.” Foot (2001), Thomson (2008), Kraut (2011), and others have put this distinction to use when discussing basic questions of value theory. In §§1-2, I outline Geach’s proposal and argue that attributive evaluation depends on a prior grasp of the *kind* of thing that is evaluated, which is another way of saying a prior grasp of a thing’s *nature*. In §§3-4, I discuss the evaluation of artifacts, which provide the clearest examples of attributive evaluation. This allows me to address a series of problems apparently facing the idea of attributive goodness. In §5, I consider the neo-Aristotelian idea that we can extend attributive accounts of goodness to human lives, and I pay attention to Foot’s account of natural goodness. This leads me to consider the goodness of human life as a whole in §6. At this point, I depart from Geach’s approach and argue that questions of attributive goodness finally give rise to questions of predicative or absolute goodness.

RÉSUMÉ :

Il y a peu de consensus sur la vérité ou les conditions de référence pour les termes évaluatifs tels que « bon » et « mauvais ». Dans son article « Good and Evil » (1956), Geach a proposé de distinguer les usages attributif et prédicatif du « bon ». Foot (2001), Thomson (2008), Kraut (2011) et d’autres considèrent qu’il faut utiliser cette distinction lorsqu’il s’agit des questions fondamentales de la théorie de la valeur. Dans les parties 1-2, je décris ici la proposition de Geach et je prétends que l’évaluation attributive dépend d’une compréhension préalable du *genre* de chose qui est évaluée, ce qui est une autre façon de dire une compréhension préalable de la *nature* d’une chose. Dans les parties 3-4, je discute de l’évaluation des artefacts, qui fournissent les exemples les plus clairs d’évaluation attributive. Cela me permet d’illustrer une série de problèmes apparemment liés à l’idée de la bonté attributive. Dans la partie 5, je considère l’idée néo-aristotélicienne voulant que nous puissions étendre notre compréhension de la bonté attributive aux vies humaines, et je me penche sur l’idée de bonté naturelle selon Foot. Cela me conduit à considérer la bonté de la vie humaine comme un tout, dans la partie 6. De là, je m’éloigne de l’approche de Geach pour affirmer que les questions de bonté attributive suscitent finalement des questions de bonté prédicative ou absolue.

1. “GOOD” AS AN ATTRIBUTIVE ADJECTIVE

Grammarians distinguish between *attributive* and *predicative* adjectives. Geach (1956) drew philosophers’ attention to this distinction and argued that it could be extended and applied to our understanding of the logical grammar of evaluative terms. Consider a complex predicate of the form “is an AN” where A is an adjective and N is a noun. For example, “is a red book,” “is a sharp spade,” or “is a good move.” By looking at the way complex predicates behave, we can draw a distinction between two types of adjectives. In some instances, the predication “X is an AN” logically decomposes into the two predications “X is an N” and “X is A.” So, using Geach’s example, “X is a red book” logically decomposes into “X is a book” and “X is red.”¹ In these cases, the truth-value of the complex predication is simply the truth-value of the conjunction of its component predications—i.e., “X is an N” and “X is A.” If it is true that X is red and it is true that X is a book, then it is true that X is a red book. When complex predicates behave in this way, we can say that the adjective is predicative.

Significantly, however, not all adjectives behave like “red.” Some complex predicates are inferentially irregular in the sense that their truth-values are not given by the conjunction of their component predicates. In particular, Geach observed that the predicates “is good” and “is bad” are inferentially irregular. Thus, the predication “X is a good book” does not logically decompose into the predications “X is a book” and “X is good.” If the proposition “X is a good book” is true, we cannot infer that “X is good” *simpliciter*. For one thing, it is obscure what it could mean to say that “X is good” (if not as an ellipsis for X is a good book).² Or, consider the proposition “X is a good poison.” We cannot make two unqualified inferences that “X is a poison” and that “X is good.” Rather we mean something like “X is good as poison” or “X has all the properties poison should have” or “X is well-fitted to doing what poison is for.” In this sense “good” and “bad” are like “big” and “small.” We cannot infer from “Fred is a big flea” that “Fred is big.” Even big fleas are rather small creatures. Rather, we should understand something like “Fred is big for a flea.”

If Geach is right that the underlying logical form of “good” and “bad” suggests that they are attributive adjectives, then the predication “X is good” is incomplete. Geach put the point as follows:

Even when ‘good’ or ‘bad’ stands by itself as a predicate, and is thus grammatically predicative, some substantive has to be understood; there is no such thing as being just good or bad, there is only being a good or bad so-and-so. (Geach 1956, p. 34)³

Although I am not persuaded that “good” is always *used* in this way—I discuss an important predicative use in §6—it seems correct that predicative uses of “good” are not normally truth-evaluable as they stand. In order to determine the truth-value of a proposition of the form “X is good,” we need to know what X is.

Moreover, depending on what X is, it may be the case that we cannot coherently employ the predicate “is good” at all. Foot (2001, p. 2) describes how she sometimes secured recognition of this point by holding up a small piece of torn paper and asking whether or not it was good. Offering to pass it round so the audience could get a better look would provoke laughter in recognition of the grammatical-cum-logical absurdity of the question. We cannot talk meaningfully about scraps of paper being good in an unqualified sense.

In a similar vein, Thomson offers the following example:

You are standing in front of the array of melons at your grocer’s, feeling helpless. Your grocer notices. He points to one in particular and says, “That one’s good.” [...] It would be utterly astonishing if when you asked, “Do you mean that that’s a good melon?” he replied, “Oh dear me no, I haven’t the faintest idea whether it’s a good melon, I meant only that it’s a good thing. (Thomson, 2008, p. 13)

Again, this is comical and the comedy comes from the fact that it generally makes no sense to say that a thing is simply good as opposed to a good instance of its kind.

Thomson (1997, 2008) argues that there is a metaphysical explanation for this sense of absurdity. It is often thought that goodness is a property. Surface grammar suggests no difference between the predicate “is good” and predicates such as “is red” and “is spherical.” As it is normally the function of predicates to denote properties, it is natural to think that a sentence such as “X is good” has the function of attributing the property of goodness to X. With this picture in mind, it is then natural to think that the primary task of ethics is to identify whatever property the predicate “is good” denotes, or the universal that particular instances of good things instantiate. Once we have done this, we can ask where this property is instantiated and how this should guide our conduct.

This view is undoubtedly attractive and a great deal of moral theory has presupposed it. Influentially, G. E. Moore held this sort of view of ethics. When he wrote that ethics is “the general enquiry into what is good,” (Moore 1903, p. 2) he meant that ethics is the general enquiry into what the property of goodness is and what has that property. Similarly, when he wrote that “the only possible reason that can justify any action is that by it the greatest possibility of what is good absolutely should be realised” (*ibid.*, p. 60), he meant that we should maximize the number of instances of the property of goodness. For this reason, Thomson (1997, p. 273) calls this sort of view “Moore’s story,” although it is far from being peculiar to Moore.⁴ Thomson also suggests that Moore’s commitment to the view that goodness is a property had a baleful influence on subsequent metaethics. For it is very difficult to see what property could be exemplified by all of the things that are good or how we should go about identifying this property. Locating the property of goodness in the natural order is an especially daunting task—all the more so for those who think that recognizing the good

necessarily motivates us. Consequently, Thomson (2008, p. 11) suggests that the property of goodness is epistemologically and metaphysically “dark”. This is reminiscent of Mackie’s complaint that moral properties would have to be “queer” for moral predicates to refer.

Moore’s own view, of course, was that goodness is an unanalyzable, non-natural property that is outside the causal order and revealed by a *sui generis* cognitive faculty of moral intuition. A more common response has been the non-cognitivist line that what is shared by all of those things we call “good” is not some mysterious property of goodness, but the fact that we approve of them in some way. Thus, normativity is built into judgment rather than the world. If, however, Geach was right that “good” and “bad” are attributive adjectives, then it appears that Moore’s story rests on a mistake. The fact that we are wont to say “X is good” disguises the fact that “is good” is an incomplete predicate. This sends us off in fruitless pursuit of the property denoted by the predicate “is good,” such as the universal *goodness* that all good things instantiate. However, it is no more plausible to think that there is a universal of goodness that all good things instantiate than to think that there is a universal of bigness instantiated by big fleas, big noses, and big attics.⁵

Saying that there is no property of goodness is neither to say that the term “good” is meaningless nor to say that there is no property, or set of properties, denoted by particular attributive uses of the term. Nor, again, is it to say that there is nothing in virtue of which propositions including the term “good” are true. Nor, finally, is it to concede that the term “good” involves us in a hopeless hotch-potch of equivocation. We are still free to investigate how attributive uses of “good” function and under what conditions it is true to say that something is good (just as there are truth conditions for attributing “bigness” to fleas, noses, and attics). But in order to identify truthmakers for propositions with evaluative content, we need to consider the *kind* of thing that is claimed to be good or bad. I approach this task below, starting with the (relatively) straightforward case of artifacts, before looking at natural functions and then human lives. I offer the qualification in advance, however, that it would be unwise to assume that there is single, legitimate way in which “good” functions in English, let alone a single, legitimate way in which cognate and near cognates of “good” function in other languages.⁶ In the end, however, my concern is much less with logical grammar or with natural language than with hints about the metaphysics of value.

2. GOODNESS, GRADING, AND KINDS

One way in which we might understand how the term “good” functions is suggested by Urmson’s (1950, 1968) observation that “good” is used primarily as a term for grading. To say that something is good is normally to say that it measures up more or less well to some standard or set of criteria.⁷ For instance, to say that an apple is good is to say, *inter alia*, that it has a pleasing taste, is pest-free, is relatively blemish-free, and stores well. To the extent that an apple meets these criteria, it is a good apple. To the extent that it falls short of these standards,

it is a bad apple. Approximation to a standard also helps us to make comparative judgments. Thus, Urmson pays careful attention to the classificatory scheme of “super,” “extra fancy,” “fancy,” and “domestic” apple grades, detailing the particular characteristics of this ordered set of adjectives (Urmson 1950, p. 151-154). In this way, evaluative judgments appear to contrast with deontological judgments. Because “good” is a scalar adjective, we can say that one thing is better, more excellent, or closer to the ideal than another. By contrast, we do not normally say that something is “righter” or more obligatory than something else.

Because “good,” unlike “right,” has comparative and superlative forms, Evans objects that Urmson’s account cannot deal with a single specimen of a kind. Evans writes:

We could not compare it with anything and, therefore, we could not say it was good in the sense of being good of its kind. There can be no superlative judgments unless there can also be comparative judgments. (Evans, 1962, p. 30)

This shows an interesting misunderstanding. When we grade an example of a kind, the comparison is not with other members of the kind—although such comparisons can guide our judgment—but with a standard or set of criteria given by the nature of the kind.⁸ Although we can grade apples, students, and many other things in order to rank them, grading is not essentially concerned with intraclass comparisons. It could be that *all* members of a kind are equally excellent, or far from the ideal. Perhaps Evans’s objection is due to the thought that we would struggle to evaluate a single instance of a kind we had not previously encountered. This is comparable to the difficulty we might face in evaluating an antique artifact with unknown origins and purpose. It would not help, however, to encounter a trove of antique artifacts with minor differences. We do not need a comparison class, but an understanding of what constitutes the relevant standard of evaluation—i.e., what kind of thing we have in front of us.

How, though, are we to determine the relevant standard or standards? The considerations of the previous section suggest that there will be no single answer. Rather, it will depend on the *kind* of thing we are evaluating. This was part of Geach’s point in saying that a substantive must be understood when we make attributions of goodness. The proposition “X is good” is typically elliptical for the proposition “X is a good K,” where K is the kind to which X belongs.⁹ As the properties of good apples are quite different from those of good students, the two kinds are not evaluated against the same standards or criteria. This suggests a two-stage procedure for determining the truth-values of propositions of the form “X is good.” First, we identify the kind K to which X belongs. Second, we measure X against the standards that are appropriate for things of kind K.¹⁰ We do not, as Moore’s story might suggest, independently determine whether X belongs to K and whether X instantiates the property of goodness.

Although these two steps of identifying and evaluating X are conceptually separable, they often rely on one another. In particular, identifying X as belonging to K can depend on identifying certain *normative* conditions that X satisfies. For instance, part of what it is to identify an object as a spade is to identify it as a digging implement. This is to say that it must have certain properties that fit it for its purpose.¹¹ “Spade” is a success term in the sense that an object that is not at all suitable for digging could not count as one. Although there are such things as rusty, blunt, and lightweight spades, there is no such thing as a spade made out of room-temperature tofu or mathematical formulae. Of course, the line between a bad spade and a non-spade may be vague or otherwise hard to determine. We can imagine, for example, a good spade that is left to rust and gradually deteriorate. Finally, it will become more rust than spade and so bad that it will no longer be a spade.¹² Sadly the same is true of us. Deteriorating function eventually undermines personhood. There are genuine difficulties here concerning vagueness. However, whatever line we take, we should not conclude that there are no persons or no spades.

3. FUNCTIONAL GOODNESS AND ARTIFACTS

I want to raise three potential concerns in this section. One concern facing kind-dependent standards of evaluation is that many kinds seem to have no plausible standards against which they can be evaluated. For instance, Raz argues:

[R]egarding many kinds of things it does not make sense to ask what is a good or well-functioning thing of that kind. There are no good or well-functioning stones, or pebbles, or streams, or hail, or snow, or mountains, or stars, or black holes, or electrons, or photons, and so on. (Raz 2003, p. 142).

A second concern is that a kind-dependent account of goodness does not yet tell us how to determine the *particular* criteria against which to evaluate an instance of a kind. Thus, even if we can establish that there are standards for a kind, we still need to determine what those standards are. A third concern is that any such standards are not properly normative. Investigating the first concern will help us to make some progress with the second. I shall postpone discussion of the third concern until the final section of the paper.

Let us take mud as an example. Mud can be good *for the purposes of*, or good *from the point of view of*, farmers, brick-makers, and bathing hippopotamuses. It is hard, however, to see what it would mean to say that a particular clod is quite simply good.¹³ Put another way, there is nothing that it is to be good *qua* mud, or for mud to have its goodness within itself, even though mud may be good *qua* something else (building material, cooling lubricant, etc.). So, it appears that some kinds, such as spades, come with evaluative standards built into their identity conditions. As Thomson (2008, p. 21) puts it, some kinds are “goodness-fixing”. Or, emphasizing the other side of the same relationship, Raz (2003, p. 39) talks of “kind-constituting values”. Other kinds, like mud, are

neither goodness-fixing nor constituted by values. Ideally, we should like a principled way to distinguish the one from the other.

Philosophers attracted to a neo-Aristotelian naturalism about value may be inclined to approach this question via Aristotle's function (*ergon*) argument. The Greek term "*ergon*" is ambiguous in a potentially suggestive and helpful way, having connotations of both "function" and "characteristic work." As Aristotle writes, "What a thing is is always determined by its function: a thing really is itself when it can perform its function; an eye, for instance when it can see" (*Meteorologica* 390a10-2). So, the ideas of a thing's function, and what it is that makes it the kind of thing that it is, are related to the idea of its *ergon*. To know, or to understand, what some kinds of things *are*, is, in part, to know what they are *for*; and this has significance when we come to evaluate instances of a kind. As I suggested above in the case of a spade, identifying something as an instance of a kind can depend on recognizing that it is at least minimally suited for performing its function. It is for this reason that the two steps of identifying the kind to which a thing belongs, and evaluating it, can be inseparable.

So, for instance, a spade, unlike the mud it disturbs, is *for* something. Because spades are for digging, a good spade has properties that fit it for that end. Because mud does not have a function, it does not make sense to think of mud as being good or bad except in relation to something else. Mud may be *used for* something, and may be good for that purpose in the sense of being useful, but it is not *for* anything. If this explanation of goodness-fixing is along the right lines, and if goodness is related to grading in something like the way I suggested in §2, then we should expect that the goodness of X will depend on the degree to which it fulfils the function of the kind K to which it belongs. This allows us to distinguish certain kinds that have built-in or intrinsic standards of evaluation. It also moves us in the direction of understanding what the standards of evaluation are for those kinds.

4. ARTIFACTS AND DESIGN

It might be objected that the example of a spade allows me to make the case too easily that the goodness of a thing depends on its function. Spades are unusual because they are artifacts, which have been *designed* for a purpose. Perhaps, then, if a spade has a function, and its function fixes its goodness as an instance of a kind, this is only because of its designer's intentions. Following this, we can distinguish two types of concern: noting that a spade has the function of digging (i) fails to pick out any non-arbitrary metaphysical truths about goodness-fixing kinds, or (ii) fails to yield any significant insights concerning the evaluation of non-artifacts. I shall pick up the second concern in the next section. According to the first concern, even if we can talk of a spade as having the function of digging, we might just as easily assign it the function of being a doorstep, a weapon, or a work of art. Moreover, even if its designer intended the spade to function as a digging implement, why suppose that the designer's intentions have any sort of priority over, say, its user's intentions? Why suppose that either

carves axiological reality at its joints? After all, a spade can be *used for* diverse purposes, including keeping doors open. It may even be good *as a* doorstop, or *make a good* doorstop. It might then seem that reflections concerning a spade's function fail to settle questions concerning its goodness. The goodness of a spade depends entirely on its instrumental value with respect to the interests it serves. A heavy spade with a sharp edge is good because it enables us to do the things we want to do—plant potatoes, bury treasure, and the like—not because it measures up to an evaluative standard given by a goodness-fixing function that is independent of its user's interests.

Some of this should be conceded. The function of a spade is evidently dependent in certain ways on human agency. It is not a coincidence that the goodness-making features of spades are, by and large, those features that give it instrumental value with respect to our needs and interests. A spade is an artifact designed with the satisfaction of our interests in mind and constructed so that it can be used by creatures like us in an environment like ours. A spade that is too heavy to lift, for example, is *ipso facto* a bad spade. However, it does not follow that we are unable to evaluate spades independently of either our *present* concerns or our *personal* concerns (which I suspect is what the worry largely amounts to). Nor does it follow that identifying an object's function is entirely arbitrary or wholly dependent on the intentions, or interests, of its designer or user. We can, for example, recognize the excellence, or shoddy workmanship, of a piece of obsolete or alien technology. We can also evaluate artifacts that we do not want—to say “this is a good spade” is not necessarily to say “I want this spade,” even a little.¹⁴ One might have absolutely no interest in planting potatoes or the like, and still correctly believe that a particular spade is a good one. So, even if artifacts are designed in response to human interests or concerns, it does not follow that evaluating the goodness or quality of the artifact must depend on taking into account the evaluator's current wants or interests. Indeed, it is arguable that we are able to evaluate artifacts in light of their function even if that function serves no *possible* human interest. Some weapons may fall into this category.

Perhaps, though, it will be suggested that it is the designer's *intentions* that matter, and that this explains the weapons case. Given that a weapon has been designed to cause terrible casualties (however contrary to anyone's actual interests), we can say that the weapon is good to the extent that it is well suited to satisfying its designer's intentions. Furthermore, it might be argued on epistemological grounds that it would be very difficult to know the function of an artifact if we were ignorant of the designer's intentions. The quality of an artifact—say, a carving—can be inscrutable when we are unaware of the role it was intended to serve, or the social norms, conventions, and connotations associated with the artifact. But it still does not follow that a designer can arbitrarily assign a function to an artifact. The reason, as I have already mentioned, is that terms that refer to functional kinds, such as “spade,” are success terms. This comes from the fact that the function of an object depends on the criteria of identity for objects of that kind. So, a spade could be entirely unfitted for use as a

doorstop or a weapon and still be a spade (perhaps it is the wrong shape or too heavy to wield), but it could not be entirely unfitted for digging and still be a spade. Although a designer may well have a function in mind, that designer cannot decide at will that an artifact has a particular function. An artifact's function depends on it possessing a set of properties, such as a suitable arrangement of its parts and materials, organized with respect to the achievement of some end.

5. NATURAL GOODNESS

The second type of objection I mentioned above allows that the function of artifacts such as spades provide us with legitimate examples of goodness, but claims that artifacts such as spades are atypical insofar as they are the products of rational agency. Therefore, they can hardly serve as paradigm cases of goodness.¹⁵ One possible response would be to argue that the world is the product of a Creator and therefore shot through with rational purpose or design. A functional account of goodness seems to fit well with a theistic framework and it is probably not a coincidence that a number of prominent neo-Aristotelians are theists.¹⁶ Given certain further assumptions concerning the attributes of God, this view would also seem to have just the right shape for grounding the to-be-pursued quality of goodness. Although this line of thought is attractive in the present context, it depends on theistic commitments that many philosophers, including myself, do not share. I propose, therefore, to set it aside and consider an alternative possibility.

It is often observed that even if the world is not the product of rational design, we are still able to understand much of the biological world in design terms. In particular, evolutionary theory provides an explanation for a large number of apparently teleological features in nature. While natural selection is unintentioned, it tends to produce organisms with parts and processes that are structured towards fitness-enhancing ends. The case looks particularly strong for parts or organs that contribute to complex biological systems. For instance, it seems natural to say that hearts are for pumping blood, that eyes are for seeing, that roots are for absorbing nutrients, etc. A plausible explanation for these biological functions can be given in terms of the evolutionary pressures bearing on an organism's ancestors.¹⁷ The reason that the function of the human heart is to pump blood and not, say, to make a thumping sound is that only the former made a contribution to the fitness of our ancestors.

These examples seem to fit well with a functional account of goodness. Given that the function of the eye is to see, it is very plausible that a good eye sees well. Perhaps more interesting, however, is the thought that evolution works at the level of behavioural traits. In particular, certain behavioural traits are adaptive for a species in the context of a distinctive way of life or "life-form" (Thompson 2008). For instance, climbing well is adaptive for squirrel monkeys because they primarily live on fruit that grows on trees. Or, again, signaling well is adaptive for meerkats because they live in social groups and are individually vulnerable to predators.

It looks attractive, then, to say that at least some traits can be evaluated in terms of their contribution to a species' way of life. But, of course, we are also a species of animal with a distinctive way of life. Perhaps most significantly, we are, as MacIntyre (1999) calls us, "dependent rational animals." We are social and linguistic animals, vulnerable to each other and to our environment, and capable of representing reasons to ourselves and to others. So, as with other animals, we might say that a trait is good when it is conducive to success with respect to our distinctive way of life. In the case of humans, stable traits of character, feeling, and thought are traditionally called virtues when they are good and vices when they are bad. As with other animals' traits, these valuable traits depend on both nature and nurture. We become virtuous, if at all, through a combination of inheritance, instruction, luck, practice, and experience.

All of the above is quite programmatic. Nevertheless, considerations such as these have encouraged a number of philosophers to develop neo-Aristotelian forms of ethical naturalism.¹⁸ They mean "naturalism" not in the sense of deferring ethics to the natural sciences, but in the sense of attempting to ground ethics in human nature. Perhaps Foot (2001) did most towards developing an account of *natural goodness* along these lines. The idea of natural goodness attractively connects the evaluation of human lives with our natural history and distinctive way of life, hence Geach's well-known comment that "men need virtues as bees need stings" (Geach 1977, p. 17). Presumably, we could have evolved quite differently, but we are as we are with our special strengths and vulnerabilities, and these facts matter when we come to reflect upon what counts as living well for beings like us.

It is important to recognize that although natural selection has partly made us what we are (and more fully made other animals what they are), natural goodness is not a concept drawn directly from evolutionary biology. What matters is the place of particular traits in the life of a species *as they have become*. What Thompson (2008, p. 20) calls "natural-historical judgments" concerns the present nature of a species. These have an evolutionary explanation, but the norms that have resulted from our natural history are not justified by evolutionary considerations.¹⁹ Advocates of an ethics of natural goodness can avoid the charge that they are trafficking in explanatory reasons under the guise of justificatory reasons or crudely reducing value to reproductive fitness.²⁰ What it is to be a good human depends on what it is to be a human, but not directly on the causal explanation of the nature of our species.²¹

6. REASON AND VALUE

Neo-Aristotelian ethical naturalism builds on Geach's insight that "good" is (typically) attributive. If we are interested in what it is to live well, we get off on the wrong foot by asking with Moore "What is good?" Instead, we should ask "What is it to be a good human being?" Moreover, our approach to answering this question shouldn't be radically disjunct from the way in which we evaluate other living organisms. We should study the parts, processes, and traits that are

fitted towards the ends appropriate for things of the relevant kind. To my mind, this is a promising approach to practical philosophy. Notwithstanding a number of prominent advocates, however, it remains a minority view. In this final section, I want to look at one possible source of resistance, which I shall relate back to the distinction between predicative and attributive adjectives. It is the concern that I briefly raised in §3—namely, that attributive goodness is not genuinely normative.

While advocates of neo-Aristotelian ethical naturalism emphasize the continuity between humans and other living organisms, it might seem that our differences matter more for the purposes of evaluative discourse and, especially, morality. In particular, it is distinctive of human rationality that we can represent putative sources of value to ourselves and submit them to critical scrutiny. As Foot observes, “while animals go for the good (thing) *that they see*, human beings go for *what they see as good*” (Foot 2001, p. 56). Irrespective of whether humans are unique in the way that Foot suggests, rational beings like us face a special sort of problem. Even if we agree that there are norms of natural goodness, it seems, on the face of it, to be an open question as to whether we should reflectively endorse these norms.²² As Slote (2003) observes, it is hard to deny that we are a warlike species or that aggression between subgroups of humans has been a pervasive feature of human history. Nevertheless, most of us would reject the conclusion that warlike behaviour is good *for* human beings.²³ It might be replied that aggression and war are (normally) harmful for creatures like us. We do not flourish under such conditions. Although this is right, Slote objects that it is unclear how this response can be directly grounded in natural goodness. If warfare is part of our distinctive way of life, it appears that the response employs an independent criterion in order to evaluate norms of natural goodness.

We should not rest content, then, with the thought that some trait is good in relation to our distinctive way of life. Rather, as rational beings we have the ability to question whether we should reflectively endorse the pursuit of particular goods. Perhaps the goal of leading good human lives understood along neo-Aristotelian lines should be rejected. Arguably, only values that can survive this process of rational scrutiny are candidates for being truly normative—for being standards that make claims on the thought and action of rational animals like us.

Here is one way to make sense of this concern. I suggested in §1 that “good” is not always used in its attributive sense. Geach (1956, p. 36) dismisses predicative uses of “good” as “a peculiarly philosophical use of words”. However, this looks like *ad hoc* monster-barring. Expressions like “friendship is good,” “health is good,” “pleasure is good,” and “God saw that the light was good” are common enough. What, then, is the point of such expressions? One possibility is that people intend to express their ontological commitments with respect to the properties of friendship, health, pleasure, or light. But this strikes me as implausible. In fact, it is unclear that everyday discourse commits us to any particular view of the metaphysics of evaluative properties (see further Johnston 1993). Most of us are hard pressed to explain the metaphysical pictures or assumptions that lie behind our moral judgments.

I propose instead that the predicative form is normally used to express another idea. I have in mind the sort of usage Wiggins describes as follows:

What the philosophical lexicographer recapitulates under “good”, we might say, is the history of our constant interrogation of the life that we lead and the place where we lead it, our constant interrogation of the things that concern us or might concern us or ought to concern us. (Wiggins 2009, p. 198)

Pace Geach, we sometimes use the word “good” in a predicative sense to mean the sort of thing that is intrinsically worthy of pursuit, or the sort of thing that finally stands up to interrogation. We do not mean that pleasure, friendship, and health are good of their kind, or that they are instrumentally valuable, but that they are good *things* and that it is, therefore, rational to pursue them.

It seems, then, that we can understand the kinds of worries expressed about the value of a life of natural goodness as reflecting the following question: Is a life that is good in the attributive sense a life that is worthy of pursuit? Thus, when we consider whether to reflectively endorse a way of life, our own or someone else’s, we can ask whether it is good in the no-holds-barred predicative sense. Such questions lead quickly on to general enquiries into the human condition and whether life is good or meaningful. I want to conclude with a few observations about the conditions under which they arise and how the relationship between attributive and predicative evaluations helps us to understand them.

In this context, consider Tolstoy’s account of his doubts concerning the meaning or value of his own life. In *My Confession*, he recalls that at the height of his literary success, financially secure and with a loving wife and family, the question kept pressing upon him “Why? Well, and then?” (Tolstoy 1987, p. 10) whenever he considered any prospective good. Repeatedly hitting upon this question, he reported:

I felt that what I was standing on had given way, that I had no foundation to stand on, that that which I had lived by no longer existed, and that I had nothing to live by. (Tolstoy 1987, p. 11)

There was apparently little doubt in Tolstoy’s mind that he was leading a good human life in the sense of achieving the kinds of goods that are typically available to us. Yet, he was unsure whether or not to reflectively endorse a life characterized by these goods. These are uncertainties that press themselves upon many reflective people at some point or another.

I suggest that we can understand Tolstoy’s predicament as follows: the attributive goodness of his life left its predicative goodness an open question. Let’s call the question of the predicative goodness of an attributively good life “Tolstoy’s Question.” Tolstoy partly assuaged his concerns, and alleviated his depression, by finding value in God’s design. If we could be confident that our natures had been shaped by a benevolent and powerful Creator, then we could also be confi-

dent that our natures are basically good. If so, to do well in the attributive sense would be to do well in the predicative sense too. If there is no such Creator, however, we may have less reason to be sanguine. Perhaps our nature is more like that of an OncoMouse or a broiler chicken in the sense that our parts and processes are organized with respect to ends that we would not reflectively endorse.

How might we provide a secular answer to Tolstoy's Question? One possibility is to look for an answer in terms of our interests or what we think of as being good *for* us. Thus, we might express the hope that evolutionary forces have molded us such that, unlike the constitution of a broiler chicken, our constitution is well suited to meeting our interests. Depending on one's theory of interests, it is a short step to ask whether attributive goodness satisfies our desires or preferences. This pattern of dialectic is both ancient and common. In Plato's *Republic*, Socrates attempts to persuade Thrasymachus that justice is part of flourishing. Philosophers in the social-contract tradition appeal to the idea of "enlightened self-interest." And the view that (moral) goodness benefits its possessor has been defended by some contemporary virtue ethicists.²⁴ Perhaps, then, an attributively good life is worthy of pursuit because it satisfies our interests. Let's call this the Socratic Answer.

The question of whether an attributively good life is good for us in the sense of satisfying our interests is certainly an important one. Although I am doubtful that we should accept the Socratic Answer, I cannot pursue that question here. Rather, I want to point out that the Socratic Answer leaves important elements of Tolstoy's Question unanswered. In particular, its egocentric starting point fails to take account of the variety of perspectives from which an attributively good life could be evaluated. It is not that any defence of the Socratic Answer must be substantively egocentric in the sense that it only considers self-interest. As most of us have a mixture of self- and other-regarding interests, an egocentric question can receive a partly other-regarding reply. The problem is that the Socratic Answer is formally egocentric. Thus, we might be persuaded that we are adapted to satisfy our own interests, but remain concerned that satisfying those interests is not a good thing *from some perspective that is not our own*. For the fact that we have the interests that we do is no less a fact about our nature than any other fact. The normative significance of our interests is, therefore, just as much open to interrogation as anything else.

To make the point dramatically, perhaps our nature is like that of the smallpox or Zika virus. From the point of view of the universe, perhaps we should wish for our own demise. Some environmental ethicists would remind us that such thoughts are not entirely silly as we stand at the threshold of the Anthropocene. However, the point is not that we should (or should not) adopt this view of ourselves. Instead, the point is that the matter is not settled by reflecting on our interests or preferences. As inveterate interrogators of what concerns us, we are led to question each aspect of our nature and this includes our interests and pref-

erences. We might express such thoughts by asking about the relationship between attributive goodness and predicative goodness, or what I have called Tolstoy's Question. Neither an account grounded in attributive goodness nor the Socratic Answer fully addresses such doubts.

ACKNOWLEDGEMENTS

Many thanks to Doug Campbell and Carolyn Mason for helpful discussion and suggestions.

NOTES

- ¹ Geach's example is complicated by the fact that ascriptions of colour terms such as "red" are context dependent. For instance, "X is a white wine" does not decompose into "X is white" and "X is a wine." White wine is a yellowish liquid. For more on this point see Thomson (2008, pp. 233-248).
- ² We can of course infer that "X is a book" and so the interest apparently lies with the adjective, not the noun. However, Rind and Tillinghast (2008, pp. 85-86) argue that this is a contingent feature of natural languages such as English and that it is possible to construct nouns that are inferentially irregular in the same way.
- ³ J. L. Austin employs the metaphors of "good" being "substantive-hungry," or as "crying out for substantives" (Austin 1962, pp. 68-69). In this sense, Austin suggests, "good" is similar to "real." Something of the same idea can be found in Hare (1952, p. 133). See also Hare (1957, p. 103) for a little more on the prehistory of Geach's idea.
- ⁴ Kraut (2011, pp. 10-15) traces a brief history of what he refers to as "absolute goodness" and finds similar stories in Plato, Aristotle, W. D. Ross, and Dworkin, among others. But see also his (2011, Appendix F) on Plato and Aristotle.
- ⁵ See Aristotle, *EN* 1096b23-26.
- ⁶ For more on this point, see Wiggins (2009, pp. 195-196).
- ⁷ The etymology of "normativity," from the Latin "*norma*" meaning "builder's square," also suggests a connection between the idea of a norm and the idea of measuring up.
- ⁸ Korsgaard (2003, p. 79) suggests that occasionally a single instance can define its own kind. She suggests that this may be true of a single, beautiful sunset.
- ⁹ By "kinds," I do not only mean natural kinds. Some *roles*, such as surgeon, student, or plumber, are also kinds in the relevant sense. Von Wright (1963, p. 19) proposes that the relevant notion of goodness for someone engaged in a particular role is *goodness at* or "technical" goodness.
- ¹⁰ Raz (2003, p. 45) defends a similar two-stage approach to evaluating genre-dependent kinds, such as the excellence of an opera or a romantic comedy. Raz's view differs from mine, however, insofar as he holds that if a work "is a good instance of its genre, then it is a good work absolutely, not only good of its kind."
- ¹¹ Cp. Aristotle *De Partibus Animalium* 640b36-641a3 and *De Anima* 412b21-3; 416a3-6.
- ¹² See further Korsgaard (2003, pp. 75-77).
- ¹³ Nor, because mud cannot flourish, is there any state of affairs that is *good for* mud. See Fletcher (2012) for one recent analysis of the *good for* relation.
- ¹⁴ Compare the oddness of thinking that "this tree has good roots" involves wanting those roots. For further defence of this point, see Foot (1961, pp. 57-60) or Thomson (2008, pp. 49-53).
- ¹⁵ Pigden (1990, pp. 147-153), for instance, presses this line of objection against neo-Aristotelian accounts of goodness.
- ¹⁶ The Unmoved Mover of the *Metaphysics* also plays an important and underappreciated role in Aristotle's ethical theory. When in Book X of the *Nicomachean Ethics* Aristotle makes the surprising move of advocating a life structured towards the end of intellectual activity (*theoria*) he suggests that it is "not insofar as he is a human that he will live like this, but insofar as there is something divine in him" (1177b26-7). The relationship between a flourishing life and participation in the divine is brought out most explicitly in the *Eudemian Ethics*: "If some

choice or possession of natural goods – either goods of the body or money or of friends or the other goods – will most produce the speculation of god, that is best, and that is the finest limit; but whatever, whether through deficiency or excess, hinders the service and speculation of god, is bad” (1249b17-20; see also 1145a6-11).

- ¹⁷ The *locus classicus* for this view is Wright (1973). For a useful overview of evolutionary approaches to the nature of functions, see Davies (2001).
- ¹⁸ See, for instance, Hursthouse (1999), Foot (2001), Thompson (2008), Thomson (2008) and Hacker-Wright (2009, 2012).
- ¹⁹ Compare my earlier comments on the relationship between a designer and a spade’s function.
- ²⁰ See, however, Millum (2006) for the complaint that this constitutes a move away from the relatively well-understood idea of evolutionary function, which partly motivates the idea of natural goodness in the first place.
- ²¹ See Hursthouse (2012) for one recent response to this sort of misplaced criticism.
- ²² Broome (2013, p. 12) expresses a similar basic concern when he objects that Foot’s norms of natural goodness are not “truly normative”. Foot’s response to this sort of worry is that the pursuit of the human good is the rational choice. She supplements this with the argument that there is a closer connection between happiness and natural goodness than is generally recognized and that it is rational to pursue what makes us happy (Foot 2001, p. 94-97).
- ²³ Compare Glassen (1957), who argues that Aristotle’s *ergon* argument conflates the *goodness* of man with the *good* of man. We should be rightly unwilling to pursue the project of becoming good people were this poorly correlated with our welfare or the satisfaction of our needs and interests.
- ²⁴ See, for example, Hursthouse (1999, pp. 163-191) and Bloomfield (2016).

REFERENCES

- Austin, J.L., *Sense and Sensibilia*, Oxford, Oxford University Press, 1962.
- Bloomfield, P., “Morality is necessary for happiness”, *Philosophical Studies*, Vol. 174, no. 10, 2016, pp. 1-16.
- Broome, J., *Rationality through reasoning*, Hoboken, NJ, John Wiley & Sons, 2013.
- Davies, P.S., *Norms of Nature*, Cambridge, Mass., MIT Press, 2001.
- Evans, J.L., “Grade Not”, *Philosophy*, vol. 37, no. 139, 1962, pp. 25-36.
- Fletcher, G., “The Locative Analysis of Good For Formulated and Defended”, *Journal of Ethics and Social Philosophy*, vol. 6, no. 1, 2012.
- Foot, P., “Goodness and Choice”, *Aristotelian Society Supplementary Volume*, vol. 35, 1961, pp. 45-60.
- , *Natural Goodness*, Oxford, Oxford University Press, 2001.
- Geach, P.T., “Good and Evil”, *Analysis*, vol. 17, no. 2, 1956, pp. 33-42.
- , *The Virtues*, Cambridge, Cambridge University Press, 1977.
- Glassen, P., “A fallacy in Aristotle’s argument about the good”, *Philosophical Quarterly*, vol. 7, no. 29, 1957, pp. 319-322.
- Hacker-Wright, J., “Human nature, personhood, and ethical naturalism”, *Philosophy*, vol. 84, no. 3, 2009, pp. 413-427.
- , “Ethical Naturalism and the Constitution of Agency”, *Journal of Value Inquiry*, vol. 46, no. 1, 2012, pp. 13-23.
- Hare, R.M., *The Language of Morals*, Oxford, Oxford Clarendon Press, 1952.
- , “Geach: Good and Evil”, *Analysis*, vol. 17, no. 5, 1957, pp. 103-111.
- Hursthouse, R. *On Virtue Ethics*, Oxford, Oxford University Press, 1999.
- , “Human Nature and Aristotelian Virtue Ethics”, *Royal Institute of Philosophy Supplement*, vol. 70, 2012, pp. 169-188.
- Johnston, M., “Objectivity refigured: Pragmatism without verificationism”, in J. Haldane and C. Wright (eds.), *Reality, Representation, and Projection*, Oxford, Oxford University Press, 1993, pp. 85-130.
- Korsgaard, C., “The Dependence of Value on Humanity”, in Joseph Raz (ed.), *The Practice of Value*, Oxford, Oxford University Press, 2003, pp. 63-85.

- Kraut, R., *Against Absolute Goodness*, Oxford, Oxford University Press, 2011.
- MacIntyre, A., *Dependent Rational Animals*, London, Duckworth, 1999.
- Millum, J., “Natural goodness and natural evil”, *Ratio*, vol. 19, no. 2, 2006, pp. 199–213.
- Moore, G. E., *Principia Ethica*, London, Cambridge University Press, 1966 [1903].
- Pigden, C.R., “Geach on ‘good’”, *Philosophical Quarterly*, vol. 40, no. 159, 1990, pp. 129-154.
- Raz, J., *The Practice of Value*, Oxford, Oxford University Press, 2003.
- Rind, M. and Tillinghast, L., “What is an attributive adjective?”, *Philosophy*, vol. 83, no. 1, 2008, pp. 77-88.
- Slote, M. “Review: Natural goodness”, *Mind*, vol. 112, no. 445, 2003, pp. 130-139.
- Thompson, M., *Life and Action: Elementary Structures of Practice and Practical Thought*, Cambridge, Harvard University Press, 2008.
- Thomson, J.J., “The right and the good”, *Journal of Philosophy*, vol. 94, no. 6, 1997, pp. 273-298.
- , *Normativity*, Chicago, Open Court, 2008.
- Tolstoy, L. “My confession”, in O. Hanfling (ed.), *Life and Meaning: A Reader*, Oxford, Blackwell, 1987, pp. 9-19.
- Urmson, J.O., “On grading”, *Mind*, vol. 59, no. 234, 1950, pp. 145-169.
- , *The Emotive Theory of Ethics*, London, Hutchinson, 1968.
- Von Wright, G.H., *The Varieties of Goodness*, New York, Humanities Press, 1963.
- Wiggins, D., “What is the order among the varieties of goodness? A question posed by Von wright; and a conjecture made by Aristotle”, *Philosophy*, vol. 84, no. 2, 2009, pp. 175-200.
- Wright, L. “Functions”, *Philosophical Review*, vol. 82, no. 2, 1973, pp. 139-168.

ON-CONDITIONALISM: ON THE VERGE OF A NEW METAETHICAL THEORY

TONI RØNNOW-RASMUSSEN

PROFESSOR, DEPARTMENT OF PHILOSOPHY, LUND UNIVERSITY

ABSTRACT:

This paper explores a novel metaethical theory according to which value judgments express conditional beliefs held by those who make them. Each value judgment expresses the belief that something is the case on condition that something else is the case. The paper aims to reach a better understanding of this view and to highlight some of the challenges that lie ahead. The most pressing of these revolves around the correct understanding of the nature of the relevant cognitive attitudes. It is suggested that the distinction between “dormant attitudes” and “occurrent attitudes” helps us to understand these conditional beliefs.

RÉSUMÉ :

Cet article explore une nouvelle théorie méta-éthique selon laquelle les jugements de valeur expriment les croyances conditionnelles des sujets qui les font. Chaque jugement de valeur exprimerait la croyance selon laquelle quelque chose est le cas à la condition que quelque chose d'autre le soit. L'objectif de cet article est de parvenir à une meilleure compréhension de ce point de vue et de souligner certaines de ses difficultés. La plus urgente d'entre elles concerne la compréhension adéquate de la nature des attitudes cognitives pertinentes. Nous suggérons que la distinction entre « attitudes dormantes » et « attitudes occurrentes » permettrait de mieux comprendre les croyances conditionnelles.

1. INTRODUCTION

The proposals to be outlined here depart from the idea that we do not have full introspective access to what our value claims express: we do not know whether it is a belief or a desire or some other attitude of ours that they express. Some people would disagree. They would claim that they do know what value claims express, either because they take themselves to have the relevant introspective knowledge or because, although they lack of this, they believe they are justified in asserting that value claims express, say, beliefs or desires. Of course, most of us know that when we make a value judgment we express a claim to the effect that something has a value. Although metaethics has clearly offered a number of incompatible, competing theories to explain this simple fact, there is no consensus on which of these theories is the correct or even the most plausible one. So, there are people who believe that, in fact, we are not justified in believing that we fully know what we express. This paper is about people like this—people who therefore adhere to the following idea:

- (i) We are rationally justified in believing that we do not fully know what we express in our value claims (i.e., claims to the effect that something bears a value).

Clearly (i) can be questioned. For instance, we might agree that philosophers in general and metaethicists in particular do not see eye to eye on the cognitivism/noncognitivism issue, but deny that they are the experts on metaethical issues; they do not in fact tell us much about what it is rational or irrational to believe about what we express in our value claims. Such a skeptical approach to what metaethics has accomplished would, in my view, be a gross mistake. But be that as it may, I will assume that the question “What do people who endorse (i) express in their value claims?” remains an interesting one, given how many people endorse (i) and continue to make value claims.

It should be stressed that those who actually believe (i)—and I believe this includes people who hold metaethical views, but who have not necessarily studied metaethics—might think there are stronger reasons to believe, for instance, that value claims primarily express beliefs. I certainly do so. However, whatever particular metaethical view we adopt, in light of (i) we acknowledge there are good arguments against that view that we cannot at present satisfactorily meet. So, although we might still think that our view is, on balance, the best available, we recognize that unless these counterarguments are not overturned, the core question “What do value judgments express?” remains open. People who believe something to this effect still make value claims: *de facto*, being in this state of mind does not prevent people from thinking that something is good or bad, or horrible or great, or better or worse. But it is the tentative hypothesis of this paper that these people’s belief in (i) is reflected in what they express in making value claims.¹

This paper, then, explores the viability of a novel metaethical proposal about what value claims made by the people just described convey—a proposal that promises to reflect the peculiar fact that these people do not quite seem to stand behind what they claim in natural-language terms. We should therefore look for attitudes that are in some way attended by a reservation. I shall refer to the proposal that I think explains people’s rational justification for making value claims whose full import remains obscure to themselves, and which I want to explore here as On-Conditionalism (henceforth, On-C).

It is worth saying at the outset that not everyone belongs to the group of people described above, and therefore On-C cannot be generalized to cover what all people expresses in their value claims. I trust the reader will keep this in mind. On-C is a proposal about what people who believe in (i) express (whether for the right or wrong kinds of reasons). It is these people I have in mind when I talk below about what “we” express.

In its initial form, On-C claims:

(On-C) All value claims are expressive of the cognitive state *that p on condition that q*.

Here is what On-C might tell us about a specific claim:

(On-C_a) “x is good,” asserted by a person in a situation S, expresses that x is good on condition that goodness exists.

I shall refer to the attitude, or attitudinal state, expressed by a value claim according to On-C as a *conditional belief*. And by “value claim” I have in mind a linguistic assertion—e.g., the uttering of a sentence, or the entertaining of a thought, containing at least one thin or thick value term (see Williams 1985, pp. 140–143). Here are two examples (henceforth I will not give any details about the speaker’s situation and so I will omit “asserted by a person in a situation S”):

- (1) “Pleasure is good.”
- (2) “Rescuing the child was a courageous thing to do.”

The gist of On-C is that there is no agreement (among people or among metaethicists in general) on what we express when we utter such sentences as (1) or (2). On-C suggests that, if we believe this, we should accept that when we make sincere value claims we are expressing conditional beliefs.

The expressions “value claim” and “value assertion” suffer from a well-known ambiguity. They might refer to some sentence, utterance, or conscious thought in a natural language employed by the speaker, or thinker, or they might refer to that which is expressed by the sentence, utterance, or thought in question. In

what follows, unless otherwise indicated, I shall have the first sense in mind. Quite what utterances (etc.) express is debatable. Do they express mental attitudes or the contents of these attitudes? Generally speaking, I find the second of these suggestions more plausible than the first. However, since in the present paper I am more interested in the attitude than in its content, I will sometimes speak as if value claims express conditional beliefs (when it would be more accurate to say that it is the content of these attitudes that is expressed). As long as we remain aware of this ambiguity, it should not give rise to any misunderstandings.

The word “express” has several meanings as well. Here, the relevant meaning is connected with the following idea: to say that the value claim p expresses the attitude A is to claim that there is an attitude A such that the evaluator has to be in this A -state for it to be semantically permissible for him or her to make the value claim sincerely.

2. METAETHICS

Metaethics has been dominated by cognitivism and noncognitivism. Cognitivists maintain that value claims characteristically express the beliefs of the evaluator. Noncognitivists contest this claim,² but they face a number of more-or-less serious objections and challenges. The most important is that set by the so-called Frege-Geach problem.³ Various noncognitivist and hybrid expressivist strategies (after Hare 1952) have been developed to deal with this problem, but I think it is fair to say that none is uncontroversial (e.g., see Strandberg 2015).

As I understand it, and for reasons I will come to later on, On-C is a cognitivist position. As a result, it avoids the Frege-Geach problem. There are essentially two kinds of cognitivism: one asserts that what value claims express are at least sometimes true; the other kind, known as error theories, assert that value claims—at least, those that attribute moral properties and relations—are always false (Olson 2014).⁴ Again, it seems fair to say that there is no consensus on the success of error theories, and that the same goes for each of the many realist, correctness theories that have been presented.⁵

I would like to stress that if people who express On-C claims are unjustified in believing that it is rational to disagree on metaethics, they should no longer express conditional beliefs when making value claims. Eventually a correctness or error theory, or an expressivist or hybrid view, or some other metaethical theory might make On-C obsolete. It bears repeating, On-C is *not* based on the idea that all extant metaethical views are equally good or bad. Far from it—some views face greater challenges than others, and some are better supported by argument.

Is there anything to be said for the idea that On-C coincides with the kind of fictionalist advice that error theories have prompted?

Metaethical fictionalists endorse one of two views: (1) the view that, as a matter of fact, we are pretending that there are moral facts when we express value claims, or (2) the view that, while we are not, as a matter of fact, involved in such pretense (and express genuine beliefs in our value claims), we should be, because there are no moral facts (cf. Burgess 1983). What exactly we are supposed to be expressing by these “fiction claims” is not particularly clear. It might be the kind of attitude we take toward cases of fiction. But, again, it is far from clear what kind of attitude this is. Walton (1990) describes it as a sort of pretense, as a make-believe (cf. Yablo 2002).

On-C does not say that we falsely believe there are moral facts. Nor does it say that our claims are fiction claims. According to On-C, we are not participating in a pretense when we make value claims, so there is a noteworthy difference between fictionalism and On-C. When we claim that *x* is good we do so, according to On-C, because, for instance, we believe that it is a fact that *x* is good on condition that goodness exists. We are not attempting to make something that is not the case appear true. We are not trying to make *x* into something that it is not.

In some ways this paper explores unknown territory. For this reason, I cannot deal with all of the many detailed issues it raises. In the interest of presenting and explaining the big picture, I need to focus on the forest rather than the trees. I shall do so by discussing two tasks that proponents of On-C will need to undertake eventually.

3. TWO GENERAL TASKS

The two tasks are:

(T1) On-C must include an account of how conditional beliefs should be understood in relation to bona fide examples of value claims.

(T2) On-C must supply accurate specifications of the content of the conditional beliefs that evaluations such as (1) and (2) express.

The specifications referred to in (T2) should be assessed with reference to their implications for logical, representational, and dispositional/motivational features of value claims. For instance, since the internalism/externalism issue is far from settled, it would be preferable if On-C claims were open to both internalist and externalist interpretations.⁶ Given the peculiar nature of conditional beliefs, it is not clear whether On-C claims are. In Section 7, I shall therefore point to a novel kind of internalism—a conditional kind that might be available to On-C internalists.

The success of On-C depends largely on whether we can specify in an accurate and plausible way what a value claim expresses. Here are some tentative specifications of the value claim “*x* is good”:

“x is good” expresses that x is good on condition

- (a) that goodness exists,
- (b) that x’s goodness exists,
- (c) that there are criteria for the goodness of x, or
- (d) that x’s being good can be rationally assessed (in terms of x’s goodness/conditional goodness).

Any plausible On-C specification will eventually turn out to be a quite complex belief. Arguably (albeit not necessarily always), when we make value claims our beliefs will be complex because we have some idea of what the value-makers of the value bearer are. In what follows, I shall have (a) in mind for the most part. In general, metaethicists agree that cognitivism is the default view of our moral discourse. (Noncognitivists tend to agree, but regard it as a kind of fallacy to regard the default view as conclusively supporting cognitivism.) Given that value language seems, at least, to communicate moral facts, (a) appears plausible. Furthermore, of the above suggestions, (a) expresses the simplest (least intellectualized) and most general approach to such factual communication. For these reasons, I will henceforth focus on (a).⁷

One might argue that On-C is not helpful at all. Consider again On-C_a. Unless we understand the meaning of “good” in its specifying part, On-C_a is not fully informative. True. But recall that On-C is not so much a competitor to extant metaethical theories as it is a theory about the mental state of someone who expresses the opposite of a belief without reservation: a belief, in other words, with reservation—i.e., what could be aptly characterized as a belief with an uncertainty proviso regarding the very nature of the claim being made. This clarification considerably waters down the complaint that On-C is insufficiently informative. Metaethics as we know it would not be put out of business or be made uninteresting just by the fact that On-Conditionalism is true. A supporter of On-C would probably argue that our belief in (i) does not in any important way have the effect that On-C claims it has on what we express with our value claims. It may be that what we express is precisely what one of the extant metaethical views tells us, and that we just do not know what it is that we express. In effect, such a view would be saying that a belief in (i) has no effect on what we are expressing. This I find hard to believe, but as I have already made clear (see note 1), I will not discuss it (or other alternatives to On-C) here.

An On-C-specification should help us determine the assertability conditions under which it would be appropriate to make the value claim. Consider On-C_a again. If we can ascertain what these conditions are by determining what x’s good-making features are, on condition that goodness exists, we have an informative account. There have to be such good-making features (this is a logical requirement). The real issue is rather to determine what these subvenient properties are. I am inclined to say that this is largely a substantive issue; it eventu-

ally hinges on how we justify our evaluations. Admittedly, in maintaining this I might be displaying a bias in favour of some metaethical view(s), in which case On-C would not be as available to everyone as I would like to think it is. The drawback would be that someone who thinks, for instance, that *x* is good, and who also thinks that *x*'s value-making properties are analytically determined, would not agree that the issue is substantive in nature. However, those who take this combination of views are not going to believe it is rationally fine for me to disagree with their metaethical theory anyway: they are what we might call metaethical dogmatists. On-C does not sit well with metaethical dogmatists.

Since On-C asserts that we express beliefs, one might be tempted to conclude that it is only available to already-convinced cognitivists. But this is a non sequitur. Noncognitivists may have good (albeit inconclusive) reasons for applying noncognitivism to what some people express, but there are not obliged to insist that what people who believe in (i) express is also something noncognitive. Quite the contrary. Noncognitivists have always found it hard to explain away the realism inherent in our ways of talking about value. On-C provides an alternative that is in line with the objective character of value language without committing us to value realism (or error theory). On-C should therefore be attractive to forms of noncognitivism underscored by value irrealism.

But a noncognitivist might insist that, according to On-C, we express something noncognitive by our value claims—where “we” here refers to people who believe, among other things, that noncognitivism is sufficiently flawed to be rationally questioned.

Perhaps, but there are good reasons why a noncognitivist should not insist on this. The most important reason is, of course, that, since it is unclear in what way the conditional form is open to a noncognitive “attitude interpretation,” such a proposal would be seriously problematic from the start.

Another worry is whether On-C is consistent with the idea that if people disagree over “*x* is good,” they will also be in disagreement if they understand this claim in terms of On-C. It would be undesirable to find that On-C is vulnerable to the kind of objections that semantic subjectivists (who cannot account for value disagreement) face. But this worry can be put aside. Obviously, if *a* claims that *x* is good on condition that goodness exists, and *b* claims *x* is not good on condition that goodness exists, *a* and *b* disagree with each other. In effect, there will also be some kind of disagreement between an ‘on-conditionalist’ and a person who claims that *x* is not good and also believes that goodness is, say, a *sui generis* property. They will be in disagreement over whether *x* exemplifies a property that the conditionalist ascribes to *x* on condition that there is goodness. Furthermore, the adversaries will in effect be locked in metaethical dispute over the existence of such properties.

4. CONDITIONAL BELIEFS

Conditional beliefs have been discussed in the burgeoning literature on conditionals—and in connected work in philosophy of mind and decision theory. Their role in metaethics has barely been explored, though. To proceed, one might try to first determine the truth-conditions of On-C_a. On-C beliefs have something to do with conditionals, so we might ask, What are the truth conditions for evaluative conditionals?

Of course, it is possible that our On-C belief has truth conditions of the kind characterizing beliefs in material implication (Cf. Grice 1989, and more recently, Jackson 1998). That would mean that to believe that B on condition that A would be to believe the proposition “if A, then B.” But, although I am still uncertain about the precise nature of the On-C attitude, I am inclined to think that what I associate with an On-C attitude is not a material implication. If it were, it would follow, in the case of On-C_a (among other things), that x is good on condition that goodness exists if and only if it is either false that goodness exists or x is good. But this equivalence claim does not capture what I have in mind. Some element in the relation between the antecedent and the consequent is lost when we turn the truth conditions of the conditional belief that x is good on condition that goodness exists into those of a material implication. Since the falsity of the antecedent is sufficient for the truth of the conditional (if we understand it as expressing a material implication), just about anything will come out as good if On-C is about a material implication and the first disjunct is false. Thus, if goodness does not exist, it will follow from On-C_a that “if goodness exists, x is good” is true. Intuitively, there must be more to the proposal than this. But whether or not our On-C belief has the truth conditions of a material implication, to avoid being obliged logically to endorse absurd conditionals, we would be well advised (in light of Gricean conversational implicatures, for example) to try first to determine what would warrant the On-C belief. In other words, we need first to ascertain the assertability of such beliefs. In what follows, I will therefore leave open the question of what, if anything, the truth conditions are. Instead I shall list some suggestions about the nature of On-C beliefs, and discuss two of these in more detail.

It seems natural to begin with the idea that to believe that p on condition that q amounts to having a belief *in* a conditional—specifically, in “if q, then p.” On this approach,

(I) a conditional belief is (nothing but) a *belief in* a conditional.

Since there is more than one kind of conditional, (I) needs to be made more precise. I will turn to this need in a moment. Meanwhile, it might be a good idea to list the alternative that I will also consider in due course:

(II) a conditional belief is (not a belief in a conditional but) a *sui generis* belief.

There are other possibilities. For instance, you might agree that an On-C belief is a *sui generis* attitude but want to resist the idea that it is a belief. You might, for example think that in order to qualify as a belief the attitude has to be truth apt and you might conclude, because you think conditional beliefs cannot be true and false, that they cannot be genuine beliefs. Instead you could endorse the notion that a conditional belief is a *sui generis* cognitive attitude but refuse to accept that it is a belief (and hence say it is an example of neither I nor II).

Limited space and the fact that any discussion of this latter example is bound to overlap with a discussion of (II) mean that I shall not discuss it here. I shall concentrate instead on (I) and (II), and eventually I will suggest that neither is satisfactory. However, as I shall argue in Sections 6 and 7, this does not leave us completely in the dark: further considerations will help us to get a better understanding of the nature of On-C beliefs.

5. A BELIEF IN A CONDITIONAL

In the literature on conditionals there is a much-discussed test for examining the conditions under which conditionals should be believed. This so-called Ramsey test is standardly applied to beliefs in indicative conditionals,⁸ but the test can be used as a guide to how we should understand other kinds of conditionals as well.⁹ Roughly speaking, it says that the conditional proposition “If *p*, then *q*” is acceptable (or accepted) in a given state of belief *S* just in case *q* should be accepted if *S* were revised with the new information *p*.

Just what this revised belief state amounts to is not obvious. Minimally, the revision would amount to giving up beliefs that are contradicted by *p*, leaving everything (or nearly everything else) intact in *S*. So, for example, to determine whether “If it rains in Berlin, I will believe there are wild tigers in Germany” is credible, I should ask myself, What if I were to believe that it was raining in Berlin? Should I accept that there are wild tigers in Germany? It seems not, so this is a conditional I should not accept. When we express an acceptable indicative conditional, we seem to be giving voice to our disposition to change our beliefs in the light of new information.

Ramsey’s test gives us some indication of what characterizes On-C beliefs.¹⁰ We could test whether a certain On-C specification expresses a Ramsey conditional (i.e., one that is acceptable as outlined above). Suppose I endorse On-C_a and that I then learn that goodness exists. What should we expect in such a case? I think it is safe to say that when I have revised my information about the existence of goodness, I should now think that *x* is good (on pain of being illogical).¹¹

But whether it is a belief in an indicative or a subjunctive conditional, it seems that it should be something that would stand the Ramsey test. So, if you accept On-C_a, the following should be acceptable:

(Pos) You ought to believe that x is good, if you learn that goodness exists.

But what about the following idea?

(Neg) You ought not to believe that x is good on condition that goodness exists, if you learn that goodness does not exist.

While I am ready to endorse **Pos** (supposing, that is, that On-C_a is correct), I am less confident about **Neg**. It seems I can go on believing that x is good on condition that goodness exists even if I believe that goodness does not exist. Of course, it is perfectly understandable that others may become upset with me if I withhold the fact that, actually, I also believe that goodness does not exist. But this has more to do with Gricean conversational implicature than it does with what we can or cannot believe.

Consider an analogous case. I might believe that x is a mermaid on condition that mermaids exist. The x with which I am confronted satisfies all the conditions that I believe a mermaid logically must satisfy to be a mermaid: it has a fish tail, lives in the sea, has the upper body of a woman, and so forth. But, although I am unaware of this, x , or at least x qua mermaid, is not a real being (e.g., because x is a hologram). Suppose I then come to believe that mermaids do not exist. If this belief does not upset other beliefs I have about what something has to be in order to be a mermaid, it seems that I could still believe that x is a mermaid on condition that mermaids exist. After all, I am saying not that x is a mermaid, but that x is a mermaid on condition that such creatures exist.

This seems to be a case in which we endorse a conditional about something where it is a merely contingent fact that this something does not exist. As far as I know, mermaids could exist. My initial inclination to accept that x is good on condition that goodness exists, and to maintain this belief even if I come to know that goodness does not exist, is perhaps likewise dependent on the belief that goodness could exist. The difference is, I suppose, that I am convinced that mermaids do not actually exist (it is a topic about which I am presently not willing to agree to disagree with someone who thinks they exist), but I am not convinced that goodness does not exist.

Prima facie, it is somewhat strange to think that goodness exists qua possibility but not in the actual world. This suggests that I have a reason to give up my intuition that I could still believe that x is good on condition that goodness exists even when I am sure that goodness does not exist. Perhaps it is simply too odd to imagine that goodness exists in a possible world but not in the actual world—which would suggest that we should give up talking about the existence of goodness in the first place. This would be fine, I think. The On-C pattern is not tied to the existence proposal (a), which is discussed here in part because it helps us to unravel some of the issues that On-C runs into. But, be that as it may, there might be a way of modifying our first response to this issue and rendering it slightly less odd.

Goodness claims are supervenience claims, so we might interpret the claim “Goodness does not exist” as entailing “(Presently) there are no (things such that they have or are) goodness-makers.” This claim strikes me as too implausible to be accepted, but it is somewhat less implausible than “Goodness does not (necessarily) exist.” Hence it might be possible to explain how one could go on with the conditional belief “x is good on condition that goodness exists” even if one happens to believe that goodness does not exist. For one could persist in the conditional belief if one were ready to believe that x might after all turn out to be good if goodness existed; and one could believe this because one thinks that if goodness existed, the properties of x would in effect be successful goodness-makers.¹² One could have thought that certain properties (a fish tail, etc.) made x into a mermaid on condition that mermaids exist. Similarly, one would think that x being generous, helpful, and so on, are goodness-makers on condition that goodness exists. The fact that goodness does not exist is no reason to retract the conditional belief.

So far it is hard to draw any conclusions at all from what has been said above. Perhaps the most we can conclude is that the idea cannot be excluded that a conditional belief is a belief *in a* conditional that in some cases might meet a Ramsey conditional. That is not much to go on, but it suggests a direction future work on On-C might take. However, as I shall argue later on, there is more to On-C beliefs than the idea that they are Ramsey-conditional beliefs. Meanwhile, it will be helpful to consider a proposal to the effect that a conditional belief is a *sui generis* attitude.

6. A SUI GENERIS ATTITUDE

Leitgeb (2007) argues that we should resist the idea that a conditional belief is (nothing but) a belief that entertains a conditional proposition. However, his conclusion is different from others who also have questioned the equation. Leitgeb does not want to give up the idea that conditional beliefs are in fact beliefs. His suggestion is that they are merely beliefs of a special kind that do not have a single propositional content.

Leitgeb takes insights from the logical systems for conditionals, probability, and belief-revision to support the idea that we should regard conditional beliefs as mental states amounting to a special kind of belief.¹³ These mental states, he thinks, are conditional, even if, as he believes, they lack conditional content (in the sense that they are not beliefs in a conditional). Leitgeb depicts these conditional beliefs as cognitive attitudes that are “neither true nor false and which have two propositional contents rather than just one” (Leitgeb 2007, p. 116).

The idea that conditional beliefs are not beliefs in a proposition is a recurrent theme in Leitgeb’s paper. As he puts it, “it will be an integral part of our theory to claim that a conditional belief is not a belief in a unique proposition of whatever sort” (Leitgeb 2007, p. 118).

Leitgeb's account is intriguing. Not all of its details are clear to me, though. For instance, I am not quite sure whether he actually identifies a conditional belief with the above dispositions or whether instead he takes conditional beliefs to be attitudes that result from, or necessarily involve in some way, a disposition to believe that B on condition that A is believed (Leitgeb uses the variables A and B). In light of the fact that what we often express when we make a value claim is an occurrent thought (with a certain natural-language content), I shall suggest in Section 7 that we should distance ourselves, when advocating On-C, from the idea that conditional beliefs are identical with the dispositional state Leitgeb has in mind. But before this, I want to set out an intuitive argument for the claim that there is more to conditional beliefs than Leitgeb's idea that a "B on condition that A" belief consists in a "disposition of acquiring the belief in B given the circumstances that A is believed" (Leitgeb 2007, p. 124).

The Ramsey-conditional dispositional state on which Leitgeb focuses fits well the description of a belief that we would have (but do not occurrently have) if certain conditions obtained. But beliefs of the kind I am interested in are different: they are occurrent beliefs that I have, but which are conditional in nature. Intuitively, there is a distinction between these and the dispositional state that Leitgeb describes—for the following reason. Suppose I think (α) that the ice cream melts on condition that the temperature does not fall below 0°C . If I occurrently believe what I think, we should expect that I will take precautions depending on whether I want the ice cream to melt or not. It seems clear that we do indeed take such precautions that involve such thoughts. But suppose next that I am in Leitgeb's dispositional state (β). If the temperature does not fall below 0°C , I shall believe that the ice cream melts. In that case, things are quite different. As for being in the dispositional (β) state, I would only take some relevant precautions if I occurrently believed either that the temperature was below 0°C or that it was above 0°C . So, if we think it is credible to say that we might take different kinds of precaution because we occurrently think (and I shall assume) believe that something is the case, on condition that something is not the case, then we are moving beyond a pure Ramsey-conditional dispositional view.

In the next section, I shall tentatively suggest that the further element the On-C account requires (over and above a purely dispositional element) is not an account of beliefs that we would have if a certain condition obtained, but rather an account of conditional beliefs as I have outlined them here. Such beliefs might have to be understood in terms of some dispositional states. In fact, they might turn out to have something in common with what Leitgeb has described. But I shall not try to clarify what such a state would involve. I think I am in good company (for very much the same reasons that Leitgeb has suggested) when I say that we presently lack the theoretical tools to describe the content of these conditional beliefs, and their dispositional bases, fully.

7. WHAT WE MIGHT EXPRESS BY “X IS GOOD”

To explain why On-C requires something more than a belief that we would have if some condition were met, we need to make a few distinctions that were brought into play in the above outline of conditional beliefs. It is customary to distinguish between occurrent and dormant attitudes (e.g., Audi 1972), and thus to separate *occurrent* and *dormant* believing. Arguably, occurrently believing that *p* is a manifestation of the dispositional state we are in when we dormantly believe that *p*. I shall also assume that a dormant belief is an unmanifested dispositional state (at least, as regards the corresponding occurrent belief), and that the corresponding occurrent belief is not necessarily the only manifestation of this dispositional state.

Consider again On-C_a and what follows if we accept the above outline. Accepting the outline, we see that the belief we express according to On-C_a is arguably either an occurrent conditional belief that is the manifestation of a dispositional state or indirectly related to this state in some relevant (causal) way.¹⁴ From a representational point of view, acknowledging this is consistent with acknowledging that the intentional content of the attitude state in which we are when we think *x* is good is in effect a conditional content of the kind “B on condition that A.” Applied to the case of “*a* is good,” this just is a way of saying that what we express in asserting “*x* is good” is a belief in a conditional that is directly or indirectly caused by some dispositional state the nature of which we do not presently fully grasp.

A caveat is in place at this point: what I am considering here is *not* the denial of Leitgeb’s claim that a “conditional belief is not a belief in a unique proposition of whatever sort” (Leitgeb 2007, p. 118). The belief that I suggested might have the content “B on condition that A” should not be identified with the conditional belief to which Leitgeb refers. It might be intimately related to it, but it is not identical with it.

Thus, we might agree that a belief qua dormant dispositional state does not have a single propositional content. Let us refer to this dispositional state as DS. However, we might also think that when we do express something, such as that *x* is good, this expression must be the product of a manifest occurrent dispositional state.¹⁵ Let us refer to this occurrent state as ODS. Arguably, there is more than one way for a DS to turn into an ODS. I shall therefore assume that it makes sense to suppose that the ODS can either be directly caused (or constituted) by the dormant dispositional state DS or be an occurrent dispositional state indirectly caused or constituted by DS.

If we acknowledge this, we can admit that the belief that B on condition that A does not fully capture the dispositional state (whatever that means) of an On-Conditional belief. Consistently with this, it might still (indirectly or directly) be the product of this state.¹⁶ Given this, we should perhaps accept that the intentional content of the attitudinal state in which we are when we think *x* is good

is in effect a conditional content of the kind “B on condition that A.” The price of this is high, though. What we express does not fully disclose our dispositional state.

Leitgeb might be right in thinking that there is a dispositional state of the sort he describes, one that allows us to discern a determinate kind of mental state—or, what he refers to as a *conditional belief*. He might even be right that no single belief in a conditional can be said to correspond to this state, if we take this to mean that the thought fully discloses what we are disposed to do when we are in this kind of state. But I am not quite sure why anyone would deny the latter. At least, it does not seem far-fetched to think that a thought with a single conditional content might be the product of such a dispositional state.

The second possibility I would like to briefly address is perhaps more plausible. What I have in mind is an alternative that takes the content of the belief to be categorical rather than conditional. So, what we express in making value claims is, in fact, attitudes with a categorical (i.e., nonconditional) content. This has the important advantage that it helps explain why people continue to make unconditional value claims even when they express conditional beliefs.

This option might appear counterintuitive, but appearances can be deceiving. If we keep in mind the distinction between the content of a belief and what this belief consists in (i.e., a certain dispositional state), and if we then add the second idea that the content of the belief need not fully reflect the full dispositional state, this alternative might be more interesting than the option I discussed above. It seems at least conceivable that a categorical belief might be the (direct or indirect) product of a Leitgebian (or of some other here unspecified) conditional dispositional state.

Actually, there is an indication that a belief with categorical content might be an expression of a conditional dispositional state. Let us suppose that I utter “x is good.” My claim might then be an expression of what could be appropriately referred to as a conditional dispositional state just in case *it is not causally dependent on the belief that goodness exists*. If it turns out that people who express the belief that x is good do not give up their belief that x is good when they learn that goodness does not exist (which is what we should expect if they had expressed a nonconditional belief), but rather modify their belief by conditionalizing it, then it seems to follow that they were expressing a conditional attitude. Hence what we expressed by “x is good” was in effect elliptical: we believe x is good but good on condition that goodness exists. (For what it is worth, in my experience this kind of response is not that uncommon.)

In fact, we might even go a step further and suggest that what makes the belief that x is good in some cases into a conditional belief is simply that it is not dependent on certain beliefs, period. There is no need to bring in any reference to a dispositional state like Leitgeb’s. What makes it into a conditional belief is (somewhat paradoxically) that it is causally resilient, at least when it comes to

some kinds of belief. This suggests that we have two hallmarks of conditional beliefs: they play an explanatory role in precautionary contexts and they are causally resilient.

Both the dispositional interpretation and the nondispositional interpretation suggest, then, how the thought that *x* is good might in a sense be an expression of a conditional attitude. But the nondispositional interpretation seems to be the more daring (and perhaps daunting) suggestion, since it implies that there are at least two scenarios where we express a thought to the effect “that *x* is good.” We might give up the belief that *x* is good when we realize that there is no goodness. If this happens, we may conclude that our original belief was (causally) dependent on some other attitudes. Most of our attitudes are dependent on other attitudes. But (a belief directly or indirectly caused or constituted by) an On-C belief seems to be dependent on beliefs that differ from those on which the categorical belief that “*x* is good” depends. So, what makes expressing “*x* is good” into the expression of a *sui generis* conditional attitude is the fact that it is a causally independent attitude. Learning that goodness does not exist will not always lead to the same result when a person has categorically endorsed that *x* is good.¹⁷

Much work remains to be done if we are to clarify the two options. Although they are clearly tentative proposals, they point in interesting directions, and this bodes well for the future of the On-C project.

It is worth highlighting the following intriguing metaethical implication of the second option. Metaethical internalists maintain that if you express the thought that *x* is good, you favour *x* in some way (on pain of being insincere, expressing an ‘inverted-commas’ judgment, or giving in to weakness of will). This is hard to account for if one is not a noncognitivist. But even cognitivists who think they can be internalists about “good” face a genuine challenge if they think “*x* is good” expresses a belief in a conditional. Since “good” is embedded in the conditional, it is not obvious how anyone can account for the motivating feature of “good.” That is, it is not at all obvious why you would be motivated to realize or promote *x* in virtue merely of believing that *x* is good if goodness exists. What the above discussion suggests is that the On-C approach does not need to add any further burden to the internalist; if cognitivism is consistent with internalism, then there is hope that even an On-C approach can be an internalist view.

It should be clear that what we have, then, is a novel version of internalism. Plausibly, even if “*x* is good” is motivating despite being an expression of a conditional belief (i.e., a resilient belief that survives in the presence of certain other beliefs), it is not obvious that the motivation would withstand certain beliefs. We might be motivated at *t*¹ by the thought “*x* is good,” but this motivation might not withstand the new belief at *t*² that goodness does not exist. This new belief might not mean that we give up on On-C_a, but it seems reasonable to suppose that we are not (or not as) motivated by our conditional belief at *t*². Thus, our motivation becomes conditionalized. But that it is conditionalized does not mean that externalism is (now) the correct position; the belief still moti-

vates at t^1 . Externalism tells us that motivation only occurs when something external is added to what we judge to be good. By contrast, On-Conditional internalism tells us that our judgment is motivating, but that it ceases to motivate us in certain cases when something external is added.

ACKNOWLEDGEMENTS

First, I wish to thank Wlodek Rabinowicz who was of great help when I began forming my thoughts on the subject. My discussions with him have been a steady source of inspiration. Two referees supplied great comments and challenged me on several matters, making me fully aware that there is much unfinished business to attend to. I would also like to thank Xander Selene for valuable suggestions. Earlier versions of this paper has been presented to audiences at Humboldt University, Lund University, and University of Tartu, and I am in particular indebted to David Alm, Henrik Andersson, Ben Bramble, Dan Egonsson, Seyyed M. Eslami, Frits Gåvertsson, Jan Gertken, Benjamin Kiesewetter, Francesco Orsi, Björn Petersson, and Thomas Schmidt for useful comments and discussion. Thanks also to Peter Gärdenfors for a most helpful discussion.

NOTES

- ¹ The idea that what we express by our value claims somehow reflects our beliefs in (i) can be questioned (I thank an anonymous referee for insights here). E.g., we might argue that when we claim “x is good” what is conveyed by this is the (belief) that x is good, period. The fact that *ex hypothesi* I also hold other beliefs about x and goodness, or beliefs whose expressed content is of a more complex kind than the straightforward belief that x is good, does not show that what I express with “x is good” is the complex belief, or any, or all, of the additional beliefs. This is certainly an understandable worry, and I do not claim to have proven this “noninclusive” idea to be false. However, even if we were to allow that it is true, it would still be of interest to figure out what kinds of other attitudes, or what kinds of complex attitude, one is in when one endorses (i) and makes value claims. To suggest that those who endorse (i) do hold conditional beliefs seems at least as plausible as to suggest that they hold unconditional beliefs. Compare two physicists: A, who believes that light is waves and particles, and B, who is convinced that light is only particles. It seems to me that for these physicists “light appeared in the room” expresses different cognitive states, and that this will eventually transpire once they realize what each of them takes to be entailed by his or her claim about light. But even if we say they are not expressing different beliefs, clarification of the beliefs they do express when they claim there is light seems interesting enough to make it worthwhile pursuing On-C as a hypothesis.
- ² “Noncognitivism” refers to a variety of theories, including various species of emotivism and expressivism. Recently, hybrid expressivism has attracted a good deal of attention. Roughly, on this view (there are a number of versions), “x is good” expresses by virtue of its meaning a belief (or a proposition) and a certain noncognitive attitude (desires, preferences). E.g., see Ridge (2007).
- ³ See Geach (1965). Compare Schroeder (2008).
- ⁴ Thus, error theorists should admit that the negation of a false moral claim expresses something true.
- ⁵ For recent discussion of some well-known challenges to moral realism, see Schechter (forthcoming).
- ⁶ According to one version of internalism, which comes without any proviso, if you endorse “x is good,” you must either be disposed or motivated to favour x in some sense (or recognize that you have a reason to favour x)—where favour is a place-holder for different kinds of positive noncognitive attitudes.

- ⁷ Limited space prevents me from discussing other candidates more thoroughly. However, since I am interested in presenting the general structure of the proposal, I believe we may confine our attention to a simple example that accords with the default position. The challenges that this example faces will in effect be a problem for all versions that follow the On-C pattern of analysis.
- ⁸ Plausibly, while some On-C specifications are best understood as involving indicative conditionals, others are to be understood as involving subjunctives. For this distinction, see Adams (1970).
- ⁹ Ramsey indicated what is now known as Ramsey's test in a famous footnote in Ramsey (1929, p. 155). The test has been used to illuminate the semantics of conditionals as well the conditions under which they should be believed.
- ¹⁰ Compare Bradley (2007).
- ¹¹ We are not entitled to conclude here that On-C_a expresses an indicative. There may be subjunctive conditionals that meet the Ramsey test, and there may also be indicative conditionals that are acceptable, but which are not Ramsey conditionals. E.g., suppose I stress how improbable something is by saying, "If he is the king of Sweden, I am the pope." If I were to learn that *x* is indeed is the king of Sweden, I would not then believe that I am the pope.
- ¹² Distinguish between the beliefs that (*) if goodness were to exist, *x* would be good; (**) if goodness exists, *x* is (contrary to what I believe, but as I think might be the case) good, and (***) if goodness exists, *x* is (contrary to what I believe and to what I think could logically be the case) good. My attitude in (*) concerns what would be the case, if goodness existed. Thus, I do not make a statement about the actual world. In (**) and (***) I do consider the actual world: I make a statement about what actually is the case on the supposition that what I believe about this world (namely, that goodness does not exist, as a matter of actual fact or by necessity, respectively) is wrong.
- ¹³ On the basis of a pair of impossibility results on conditionals, probability, and belief-revision, Leitgeb argues that conditional beliefs differ from beliefs in conditionals qua mental states. The results come from distinct research traditions: belief revision and Bayesian (probabilistic) approaches to rational belief change. Bayesians use subjective conditional probability measures to explain rational change. Lewis (1976) proved that equating the probability of a conditional "if A, B" with the conditional probability of B, given A is impossible, on pain of triviality. Gärdenfors (1986), in his turn, showed that the Ramsey test for the assertability of conditionals is problematic. This test was shown by Gärdenfors to be incompatible with the norms of rationality for belief revision if we represent belief states as sets of propositions or sentences and take conditionals to be possible members of such sets (Gärdenfors 1986; cf. Rabinowicz and Lindström 1995). This claim has been contested by Bradley (2007), who argues that we should not conclude from the impossibility results that the Ramsey test is false; it is rather what he refers to as the "preservation condition" that creates a challenge for the Ramsey test. The preservation condition states that if a person *a* learns a sentence that is consistent with *a*'s prior beliefs, *a* retains all these prior beliefs after revision. Bradley argues that it is this preservation condition, together with some other principles, that leads to unacceptable results.
- ¹⁴ We should be open to the possibility that what value claims express is directly or indirectly identified with conditional beliefs. Although the direct or indirect accounts are quite different, I shall not consider in this paper which is the more plausible.
- ¹⁵ That is, to explain the difference between occurrent and dormant attitudes, we need the distinction between a manifested dispositional state and an unmanifested dispositional state.
- ¹⁶ Among other things, it would mean, it seems, that the assertability conditions for "x is good" could be met even if it were false that *x* is good.
- ¹⁷ It might still be dependent on some beliefs—e.g., it would not resist the belief that goodness logically, or necessarily, could not exist.

REFERENCES

- Adams, E. W., "Subjunctive and indicative conditionals", *Foundations of Language*, vol. 6, no. 1, 1970, pp. 89-94.
- Audi, R., "The Concept of Believing", *Personalist*, vol. 53, no. 1, 1972, pp. 43-62.
- Bradley, R., "A Defence of the Ramsey Test", *Mind*, vol. 116, no. 461, 2007, pp. 1-21.
- Burgess J., "Why I am Not a Nominalist", *Notre Dame Journal of Formal Logic*, vol. 24, no. 1, 1983, pp. 93-105.
- Geach, P. T., "Good and Evil", *Analysis*, vol. 17, no. 2, 1956, pp. 33-42.
- , "Assertion", *Philosophical Review*, vol. 74, no 4, 1965, pp. 449-465.
- Grice, P., *Studies in the Way of Words*, Cambridge MA, Harvard University Press, 1989.
- Gärdenfors, P., "Belief revisions and the Ramsey test for conditionals", *Philosophical Review*, vol. 95, no. 1, 1986, pp. 81-93.
- Hare, R. M., *Language of Morals*, Oxford, Clarendon Press, 1952.
- Jackson, F., *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford, Clarendon Press, 1998.
- Leitgeb, H., "Beliefs in conditionals vs. conditional beliefs", *Topoi*, vol. 26, no. 1, 2007, pp. 115-132.
- Lewis, D., "Probabilities of conditionals and conditional probabilities", *Philosophical Review*, vol. 85, no. 3, 1976, pp. 297-315.
- , "Probabilities of conditionals and conditional probabilities II", *Philosophical Review*, vol. 95, no. 4, pp. 581-589.
- Lindström, S. and Rabinowicz, W., "The Ramsey Test Revisited", in G. Crocco, L. Fariñas del Cerro and A. Herzig (eds.), *Theoria*, Oxford, Oxford University Press, 1995, pp. 131-182.
- Olson J., *Moral Error Theory: History, Critique, Defence*, Oxford, Oxford University Press, 2014.
- Ramsey, F., "General Propositions and Causality", in D. Mellor F., (ed.), *Ramsey Philosophical Papers*. Cambridge, Cambridge University Press, 1929, pp. 145-163.
- Ridge M., "Ecumenical Expressivism: The Best of Both Worlds?", in Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics*, Vol. 2., Oxford, Oxford University Press, 2007.
- Schechter, J., "Explanatory Challenges in Metaethics", in T. McPherson and D. Plunkett (eds.), *Routledge Handbook of Metaethics*, London, Routledge, forthcoming.
- Schroeder, M., "What is the Frege-Geach Problem?", *Philosophy Compass*, vol. 3, no. 4, 2008, pp. 703-720.

Strandberg, C., "Options for Hybrid Expressivism", *Ethical Theory of Moral Practice*, vol. 18, no. 1, 2015, pp. 91-111.

Walton K., *Mimesis as Make-Believe*, Cambridge, MA, Harvard University Press, 1990.

Williams B., *Ethics and the Limits of Philosophy*, London, Fontana, 1985.

Yablo S., "Go Figure: A Path Through Fictionalism", *Midwest Studies in Philosophy*, vol. 25, no. 1, 2001, pp. 72-102.

TWO CONCEPTIONS OF PRACTICAL REASONS

CHRISTOPH HANISCH

ASSISTANT PROFESSOR, DEPARTMENT OF PHILOSOPHY, OHIO UNIVERSITY

ABSTRACT:

I discuss and compare Joseph Raz's and Christine Korsgaard's accounts of reasons for action. One fundamental disagreement separating the two approaches is the role that they assign to two central features of practical deliberation: Korsgaard assigns priority to identity-constituting practical principles, whereas for Raz reasons are the fundamental normative units. In the course of this comparison, two claims are defended: (1) Taking-up a realist stance vis-à-vis one's reasons is a non-optional feature of one's first-personal deliberative standpoint. This remains true even if this renders the workings of an agent's reasons analogous to a placebo. (2) The motivational aspect of an agent's reasons should not be conceived along causalist lines. Exploiting a thought by Hegel, I discuss an alternative conception of reasons' temporal genesis and force—i.e., one that allows reasons for a particular action to emerge at a later point.

RÉSUMÉ :

Cet article résume et compare les théories de Joseph Raz et de Christine Korsgaard sur les raisons qui supportent l'action. Un désaccord fondamental séparant ces deux approches tient au rôle qu'elles attribuent à deux caractéristiques essentielles de la délibération pratique : Korsgaard accorde un primat normatif à l'identité constituant les principes pratiques, alors que Raz accorde ce primat aux raisons comme unités normatives. Au fil de cette comparaison, deux revendications sont défendues : 1) L'adoption d'une position réaliste quant aux raisons de quelqu'un est une caractéristique non facultative de son point de vue délibératif à la première personne. Cela reste vrai même si cela rend le fonctionnement des motifs d'un.e agent.e analogue à celui d'un placebo. 2) L'aspect motivationnel des raisons d'un.e agent.e ne doit pas être conçu selon une perspective causaliste. Exploitant une idée de Hegel, je discute d'une autre conception possible de la genèse temporelle et de la force des raisons d'agir, ce qui permet d'envisager qu'elles peuvent aussi apparaître après coup.

Joseph Raz's and Christine Korsgaard's accounts of practical reasons are thought to stand in sharp opposition to each other (Scanlon, 2004, p. 233).¹ Raz's (1999; 2012) realist account tells us things like, "reasons are facts" to which agents (have to) respond. Reasons for actions are "aspects of the world." This Razian realism about reasons, as well as similar frameworks defended by Scanlon (2014), Parfit (2011, Vol. 1, Part 1) and many others, contrasts with the Kantian constructivism advocated by Korsgaard (2008; 2009), Onora O'Neill (1989; 1996; 2013) and Sharon Street (2008; 2011). According to constructivism, practical reasons are the result of our deliberative activities that we agents engage in, in accordance with constitutive principles of practical reasoning. And indeed, there are many moments in this debate that suggest that these are mutually exclusive positions concerning the ontological nature and practical role of reasons.

I attempt to establish some common ground between these two views and focus on the relative status that Raz and Korsgaard assign to practical reasons and practical principles, respectively. This paper's aim and scope will be limited in so far as it singles out Raz and Korsgaard as representatives of the realist and constructivist camps respectively and focuses on their work. The notions of *reasons* and *principles* are defined in sections one through three. Raz insists that reasons, broadly understood as features of actions that make them worth choosing or rejecting, have conceptual and practical priority over principles. According to Razian realism, practical principles play the mere subservient role of so-called *rules* that assist deliberators in systematizing the reasons that apply to them independently (of such principles). Korsgaard takes the opposite view. Her account of normativity asserts that the principles of choice that an agent identifies with are the source of her reasons. The principles that define an agent's self-conception are at the same time always employed when she endorses a proposed action, turning the latter into an act done for a reason.

My account consists of two suggestions, defended in sections four and five. The first one is the placebo analogy: treating one's reasons as having an agency-independent dimension along Razian lines is an attitude that agents constitutively (have to) incorporate into their self-understanding lest their own first-personal stance should end up losing its foothold in a stable normative realm. Agents generate practical reasons on the basis of their identity-constituting principles. But in order for *practical* deliberation to fulfill this function, deliberators must adopt the view that this process handles items that gain their justificatory force independently of their volitional apparatus. If agents were to render this process fully transparent to themselves and treated the assumption regarding the factual nature of reasons in a constructivist manner all the way down (a manner that is presumed to be solidly justified for the sake of this paper's exercise of combining the two views), the very self-constituting activities that Korsgaard singles out as every agent's central practical task would undermine themselves and the agents' identities would destabilize. Like a placebo that ceases to unfold its beneficial health effects upon being unmasked as a placebo, an agent's reasons would lose their force if they were conceptualized in the way that Kantian constructivism suggests. For the sake of their own agency, then, agents must maintain an

illusion and keep on living with a threatening instability that emerges between an ontological constructivism, on the one hand, and a practical realism about reasons, on the other.

The second suggestion for overcoming the impasse between realists and constructivists focuses on the relationship between deliberation (in terms of the aforementioned principles), on the one hand, and the phenomenon of acting for a reason, on the other. Raz claims that the latter can well take place, even if no deliberation whatsoever precedes the action in question. In response to Raz's examples of automatic, opaque, and swift actions, my account defends the view that our reasons should not be understood as analogous to causes regarding their temporal position relative to the actions that are chosen on their basis. Employing a thought by Hegel, I present a conception of reasons' genesis and force that allows reasons *for* a particular action to emerge later and after the non-reasons-based features of an action are already completed. This suggestion is then spelled out in more detail to accommodate the intuition that reasons seem to motivate us before the action takes place, on the one hand, and with respect to our accountability practices, on the other. Both of these issues cause problems for my deliberation-centred account of reasons.

1. RAZ ON REASONS AND PRINCIPLES

Raz defines reasons for action (I use the term interchangeably with “practical reasons”) as “facts which constitute a case for (or against) the performance of an action” (Raz, 2012, p. 36). Raz's realism about reasons becomes apparent in his discussion of the conceptual connection between reasons and normativity. He claims, with respect to all normative reasons (of which reasons for action are one sub category), that “aspects of the world are normative in as much as they or their existence constitute reasons for persons, that is, grounds which make certain beliefs, moods, emotions, intentions, or actions appropriate or inappropriate” (Raz, 1999, p. 67).

Raz assigns strict priority to reasons over principles. Practical principles are “general propositions, in our context general normative propositions, which are thought to be particularly illuminating, or central relative to a certain normative view or theory” (Raz, 2012, pp. 6-7). And with respect to such principles, which he explicitly thinks are occupying *the* foundational role in views “descending from Kant,” Raz says: “Given that what one has conclusive reasons to do [...] depends on what reasons for and against the action one has, *principles are conclusions from the existence of reasons, and cannot themselves occupy a fundamental role in an account of normativity*” (Raz, 2012, p. 7; my emphasis).

Raz spells out the case against principles in more detail. He presents a number of intuitive cases in which agents seemingly perform full-fledged actions and do so both intentionally *and* for a reason without, at the same time, acting on any principles of practical deliberation. Principles rarely figure in the agent's choice to help others: “You see children torturing a cat and you stop them. No princi-

ple figures in your decision” (Raz, 1999, p. 226). Again, rescuing the cat is an action that is intentionally performed (the agent is controlling what he or she is doing) and it is taken for a reason (the agent, if asked, would mention “prominent aspect(s) of the situation”), such as that the cat is in pain.

Raz emphasizes that intentionality and action based on a reason do not necessarily involve the active deployment of practical principles. He assigns principles the minor role of guidelines that “may point out an important consideration to be taken into account, so that being guided by such a principle is the same as giving *the reason to which it points* due weight in one’s reasoning” (*ibid.*, p. 227; my emphasis). Similarly, principles “are valid if they *represent the correct* outcome of rational deliberation about the merit of the various reasons which bear on [...] decisions” (*ibid.*, p. 227; my emphasis). These statements highlight the priority that Raz assigns to reasons over principles. Principles get employed by practical deliberators (if at all) when they face a hard choice, in which strong reasons are present in support of two or more available courses of action. In such situations “rules of thumb” (*ibid.*, p. 226) might provide assistance to struggling deliberators and guide them in acknowledging and organizing the reasons relevant to their options. In such cases, agents may take recourse to pragmatic rules or even more substantive moral principles and policies in order to identify the course of action that is best supported by the independently subsisting reasons. Still, according to Raz, the practical reasons are *there* to be recognized, and they are the fundamental building blocks of deliberation with respect to which organizing principles are then (sometimes) applied.

2. KORSGAARD ON PRINCIPLES AND REASONS

According to Korsgaard’s neo-Kantian account of action and agency, every process of practical deliberation begins with an incentive. An incentive is a perceived “good-making property of an object” of choice. Incentives are the attractive features of the objects that figure in our actions (for example, the aesthetic features of a painting that make it the proper object of us celebrating it) (Korsgaard, 2009, pp. 122-124). The Kantian approach insists that incentives in and of themselves do not provide reasons for actions, notwithstanding the fact that agents often refer to these incentives as if they were free-standing facts. In order to get to reasons, agents need to enrich incentives with another Kantian component—namely, inclinations. Inclinations are agents’ reactions to the incentives in question. They are similar to desires in that they propose certain courses of action. The controversial claim of the neo-Kantian and constructivist account is that it is ultimately we who confer value on objects of choice by being inclined towards them. Without rational agents and their inclinations, understood as reactions to attractive features of objects, there wouldn’t be any values or reasons. Inclinations are therefore a necessary condition for the occurrence of reasons for action.

However, the presence of an inclination is of course not a sufficient condition for generating reasons for action, Kantians insist. And this is where practical prin-

ciples enter the picture. They provide the cognitive means that are necessary to consciously decide whether or not to identify with an inclination—that is, whether or not to endorse the suggested course of action that the inclination proposes. Practical deliberators and agents have to employ self-imposed and choice-constraining principles to get from incentives, via inclinations, to reasons for action. In her recent work, on the function of action in the course of individual self-constitution, Korsgaard (2009, pp. 23-24) clarifies the way in which our normative self-conceptions and their principles play this role of a necessary condition in the genesis of practical reasons:

We must act, and we need reasons in order to act. And unless there are *some* principles with which we identify we will have no reasons to act. Every human being must make himself into someone in particular, in order to have reasons to act and to live. Carving out a personal identity for which we are responsible is one of the inescapable tasks of human life.

The crucial bit in this quotation is the identification of the relationship between (identity-constituting) principles, on the one hand, and reasons for action, on the other, as one of necessity (“unless”). Without at least one principle that an agent endorses and identifies with, there cannot be any reasons for this agent. Since, however, reasons are necessary for action and, furthermore, actions are the indispensable means of constituting ourselves into agents to begin with, we (qua minimally existing agents) inescapably have to identify with some principles.² “To have a reason” means that one understands oneself as an agent whose deliberative stance is structured by certain normative principles, the employment of which consists in taking up incentives in a particular way rather than another. The resulting inclination, the pulling-together of one’s deliberative stance in accordance with these deliberation-guiding principles, is, together with the initial incentive, the reason for the action that one chooses. These practical and identity-constituting principles are foundational in a twofold sense: They define and determine under which normative description we regard ourselves as agents. And, at the same time, they guide our deliberation and, ultimately, our (self-constituting) actions upon encountering objects of potential choice with their specific properties.³

One should also briefly acknowledge Korsgaard’s important claim that the apparent contingency of agents’ specific practical identities (and principles) does not result in practical reasons’ lack of normative force. On the contrary, the principles that constitute one’s deliberative stance are normatively binding, because they instantiate one way of dealing with an inescapable problem, the task of self-constitution and agency, successfully. You have to conceive of yourself as being normatively bound in one way or the other, in order to conduct your own practical reasoning in a way that is minimally intelligible (at least to yourself). Furthermore, the reasons that spring from endorsed inclinations are capable of motivating the person who has them, because these reasons are sourced in how the agent experiences herself in responding to the necessity to act. The elegance of Korsgaard’s picture is that it accounts for reasons’ normativity *and* their intrinsic

sically motivational nature by bridging the gap between independent features of the world that trigger an agent's process of practical deliberation and action, on the one hand, and the value-conferring power of human agency, on the other.

3. PRINCIPLED REASONS AND REASONED PRINCIPLES

The comparative reconstruction of two paradigmatic accounts of practical reasons has resulted in a conclusion that seems to support the widely shared diagnosis that realists and constructivists present sharply conflicting approaches. Reasons for action and practical principles occupy the opposite priority-status in the two theories. The Kantian conception of reasons—i.e., that they originate in rational agency—favours an account that prioritizes principles over reasons, in particular when the former are considered the constitutive elements of the deliberative standpoint, from which reasons are generated (as opposed to recognized). According to the Razian alternative, reasons take priority over principles exactly because the task of the latter merely is to guide the activity of organizing and systematizing “normative facts” and “aspects of the world” in the correct manner.⁴

In the remainder of this paper I evaluate these two views by means of applying them to problems in practical philosophy and action theory. That way, the strengths and weaknesses of Korsgaard and Raz will come into view more sharply, and overcoming the weaknesses of the one will naturally point towards some strength of the other. The suggestions have already been mentioned. First, understanding one's reasons along the lines of agency-independent facts is itself constitutive of first-personal deliberation and self-constitution. A benign analogy with the placebo effect shows that the presupposition of realistically conceived reasons is itself a principled precondition for getting the constructivist process of generating reasons going in the first place. Second, the inherent action-directedness of practical reasons too can be explained by unifying the two accounts. It is the act of *identifying* with certain actions that explains why the resulting reasons are necessarily capable of motivating those who have them. However, to get to this conclusion, I will need to introduce a rather unorthodox conception of how actions relate to those reasons that serve as the basis for choosing them. This conception parts company with the view that reasons always work analogously to causes, particularly concerning their temporal priority over the actions that result from endorsing them.⁵

4. THE PLACEBO ANALOGY

One of the main worries constructivist accounts of practical reasons encounter is that they seem to leave us with an unpleasant subjectivism about practical reasons (Regan, 2002). Critics of the constructivist paradigm claim that there is an arbitrariness about our reasons for action that comes with the strategy of fully internalizing their sources. If I reflectively *want* Anscombe's (2000, pp. 70-71) infamous saucer of mud for its own sake, then this inclination constitutes a robust reason for preferring to have the saucer and to take the related actions. Or think

of Parfit's (2011, p. 83) recent example of someone who, after ideal deliberation, wants to be in a future state of pain, again non-instrumentally so. And if it is only the first-personal side of the encounter with an object's features that accounts for the emergence of reasons, then an agent with the practical identity of a mud-lover successfully acts for reasons to the same extent that agents who engage in actions that are commonly regarded as more worthwhile.

Of course, Kantian constructivists have responded to this worry. Teleological varieties, for example, have recourse to normatively potent and choice-constraining conceptions of human nature and flourishing in order to block the mud-lover from generating reasons for his or her actions.⁶ Closer to Kant are accounts of rational willing that put certain formal constraints (such as the categorical imperative and the principle of instrumental rationality) on that very activity. My attempt to save the constructivist approach to practical reasons from the subjectivism objection differs from both of these strategies. It is more modest (or, some realists might say, disappointing) in that it fully absorbs the objective and external constraints on the agents' value-conferring activities into the deliberative perspective. My suggestion is that human beings cannot go about their lives as deliberating and acting agents without postulating and presupposing the independent status of their practical reasons in an objective and realist mode of reflection like the Razian one. This remains the case, notwithstanding the well-justified pull that makes agents acknowledge the dependence of their reasons' existence on their own activities of willing, deliberating, and choosing. Below, I employ an extension of one of Korsgaard's own central constitutive arguments to support this claim regarding the threat of an unstable agency that constructivism poses (if it is adopted practically and all the way down). The argument at stake is Korsgaard's attack on the possibility of what she calls "particularistic willing." This argument can be extended to cover the problematic nature of "purely constructivist willing," as I will call it.

But before I get to that I have to introduce the aforementioned analogy in order to show that an attempt at combining Raz and Korsgaard does not result in an unstable position. Take the phenomenon of how placebos are thought to work and of what, conceptually speaking, turns a substance into a placebo drug.⁷ Once patients understand that the improvement of their condition had originated *solely* in their volitional set up, and acknowledge that the assumed efficacious power of the drug had been merely chimerical, exactly then does that power start fading. Patients will regard it as unacceptably incoherent to find themselves in a cognitive state in which, on the one hand, they believe that the substance they have taken is causally completely inefficacious, but, on the other hand, continue to tell themselves that it is the drug that has improved their condition. It *constitutively* comes with the idea of an efficacious drug that it is *the drug* that is (at least partly but necessarily) responsible for the achieved effect. If this weren't so, and if the effect were considered to be producible entirely from within, patients would not be able to tell themselves that they had had any rational justification to take the apparent drug in the first place. After all, if internal mental mechanisms are acknowledged to be a sufficient means for bringing about the

desired beneficial effect, then taking what one knows to be a causally inefficacious substance becomes a redundant and irrational choice.

Now, what does all this have to do with the way practical reasons work? The placebo analogy is supposed to illustrate how certain, third-personal, assumptions concerning one's own reasons' independent standing are a non-optional presupposition for occupying the practical standpoint of agency. It is exactly in order for these reasons to exhibit first-personally perceived authority and normative force that agents (partly) conceptualize them in the form of choice-constraining standards with a realist phenomenology. In their inescapable roles as practical reasoners and agents, just like the patient who is in need of a cure, they wouldn't be able to generate these agency-sustaining effects if they *completely* negated and ignored a third-personal dimension regarding the so-called nature of their reasons—even if the account of that nature turned out, upon ontologically well-informed reflection, to be an unsustainable illusion. Part of the account defended here is that an essential part of a successfully constituted deliberative standpoint consists precisely in the ability to take a step outside of that very standpoint. This ability then manifests itself in an agent asking what the correct reasons *are* that he or she attempts to identify in the course of practical reflection and decision making. Employing exclusively constructivist resources all the way down and in a fully self-transparent manner won't get the process of establishing authoritative reasons off the ground.⁸

Let me defend this set of claims a bit more. In terms of the two theories that I am trying to combine, looking at practical reasons in the Razian, substantive realist, mode is the adequate one *for* those who are engaged in practical deliberation and reasoning. And this remains practically true even if it turns out that there are strong metaethical grounds on which this realism about practical reasons is theoretically not sustainable. In other words, as practical deliberators and agents, humans need Razian realism about the reasons they have, even if, on a different level of reflection, they might have to acknowledge them as originating in their subjective self-understanding. Part of what drives the placebo analogy is the thought that a workable self-understanding regards itself as being directed at something external to (and independent from) its own stance. It is from agents' individual and subjective perspectives that they *must* treat their reasons as if they pointed towards a realm of independently subsisting normative facts that are not entirely conditional on the operations of choice and volition. This practical realism about reasons is unusual because the realist presupposition in question is not primarily born out of any epistemological or metaphysical necessities. Rather, it is a *practically* vindicated presupposition through and through. The presupposition is justified because it has its ultimate source in the inescapability of deliberation and action and because the task of constituting one's agency (in the course of action) requires that agents regard themselves as located in a stable and predictable normative environment.⁹

In summary, the first suggestion of how one might go about establishing some common ground between realism and constructivism employs the placebo anal-

ogy in order to highlight a certain ambivalence that comes with the status as practical deliberator. Treating reasons realistically is beneficial relative to our needs as agents, and is a perspective that we are justified in shielding from the urge to render ourselves and the world as transparent to ourselves as possible. This antiphilosophical suggestion identifies a tension between the probable ontological truths about reasons, on the one hand, and the needs of agents and practical deliberators, on the other—needs whose satisfaction requires that agents abstain from practically acknowledging such truths in an authentic manner. The practical lives of agents are dependent on managing this ultimately irresolvable tension concerning the objective and subjective sides of practical reasons, the internal and external sources of normativity, and the realist and constructivist natures of those aspects of the world that are normatively relevant to these lives.

Now the placebo analogy allows me to return to the issue of whether to assign practical principles a foundational or merely a derivative role in an account of practical reasons. Recall that for Raz and other “reasons-fundamentalists” principles play a subservient role relative to practical reasons, whereas for many constructivists these reasons are inconceivable unless practical principles guide processes of practical reflection regarding the inclinations that humans encounter. It is at the level of those very principles that the placebo proposal highlights another, more specific, practical demand in line with its core claim that agents have to deliberate and choose as if they responded to independent reasons qua normative facts.

This way of looking at things gets me closer to one of Kant’s (and Kantians’) central theses, which is that willing an action’s end, conferring value on it, cannot take on just any form whatsoever. Willing has to be *rational* willing and as such it has to satisfy the minimum requirement of being of a universal form. The central idea, manifested in the most abstract version of the categorical imperative, is that the resultant practical reasons have to satisfy the requirement of having implications that go beyond the currently considered single instance of action. In order to exist as a unified person across time, an agent must conceive of himself or herself as employing principles that express the commitment that, at a different point in time, he or she would choose in the same way, were the circumstances similar enough. That’s the argument against the possibility of “particularistic willing” (Korsgaard, 2009, pp. 72-76). If, contrary to this Kantian picture, particularistic willing were an option, “it would be possible to have a reason that applies only to the case before you, and has no implications for any other case” (*ibid.*, pp. 72-73). And, “if you have a particularistic will, you are not one person, but a series, a *mere heap*, of unrelated impulses. [...] Particularistic willing lacks a subject, a person who is the cause of his actions. So particularistic willing isn’t willing at all” (*ibid.*, p. 76). The Kantians’ ambitious conclusion is that this minimally universal character of *all* practical reasons leads straightforwardly to robust principles of morality—the moral law, the formula of humanity, and the kingdom of ends.

While these ambitious conclusions of Kantian constructivism are not endorsed here at all, there is an important insight expressed in this critique of radical particularism.¹⁰ In order to be an agent, the principle-structured deliberative perspective, which each agent has to sustain in one form or another, has to consist of action-guiding principles that transform the incentives that the agent encounters into reasons and which do so in a temporally stabilizing manner. The placebo analogy adds that the principles that make up one's practical standpoint have to exhibit more than just this minimally universal form. They have to align agents in a certain direction—namely, one that conceives of the chosen objects as embedded in a stable and predictable external environment, an indispensable part of which is a coherent and non-arbitrary setting of normative facts. This source of stability has to be conceived as having a foothold in an agency-independent realm, a realm that must ultimately be presumed because the stability and coherence of one's own principle-structured and temporally-extended identity depend on it.¹¹

The Korsgaardian argument against particularistic willing, therefore, gets it right, but it is not the end of the story. It must be supplemented by a thesis against what one might call “purely constructivist willing.” Such constructivist willing would be one that is incoherent in the same way that particularistic willing is. If it were possible, purely constructivist deliberation about proposed courses of action would amount to willing that takes itself to be entirely detached from any guidance by fact- and world-directed principles of choice and action. A workable self-conception and practical identity must avoid what has been called, in a different context, “frictionless spinning in the void” (McDowell, 1996, p. 50). The purely constructivist and self-certifying deliberations of a fully self-transparent and self-aware Korsgaardian agent¹² would, so the suggestion goes, amount to such a spinning in the normative void that undermines the constitution of individual agency as a phenomenon in the world and in the company of other, independent, agents.¹³

5. THE ROLE OF REASONS IN ACTION: DELIBERATION AND MOTIVATION

With respect to the issue of the motivational force of reasons for action, internalist conceptions (Humean as well as Kantian¹⁴) seem to be in a much better position than realist and externalist varieties à la Raz and Scanlon. Recall Korsgaard's account sketched above: reasons for action motivate agents exactly because they emerge in the course of reflective and deliberative processes that inevitably matter to them. Agents have to decide what to do—in one way or another. And engaging their principle-structured practical identities in order to figure out whether or not to endorse any of the suggested courses of action is non-optionally exerting efficacious effects on their volitional apparatus. The inclinations that agents reflectively endorse on the basis of their identity-constituting principles are the ultimate bedrock of normative motivation. Practical reasons normatively motivate agents because of the way in which their genesis is always connected to features of their volitional experience that have the rele-

vant potential to get them moving—to use constructivist language. Importantly, these active and quite demanding processes of practical deliberation seem to precede the execution of the actions in question. Practical reasons have much in common with causes (regarding their temporal ordering, in particular). Agents encounter a powerful incentive and the related inclination; they activate choice constraining and identity-defining principles and the resulting reason in support of the proposed action normatively motivates these agents to engage in it (in case the principles deem the proposed action worth choosing and executing). It seems as if this picture of how reasons motivate agents easily overcomes the problems that Raz’s externalist separation between reason(s) and the will leaves us with.

There are, however, problems with the Kantian/Korsgaardian account of the relationship between reasons and motivation. I should consider one that is suggested by some of Raz’s remarks and by his critique of motivational internalism. This objection has to do with the so-called “guiding problem” that Raz (1999, pp. 230-238) discusses in his evaluation of Dancy’s particularism. The relevance of the guiding problem for the discussion of motivation becomes visible when one looks at Raz’s formulation of it. He says, “the issue turns on whether there is more to people’s reasons than those factors which figure in their deliberations, and which they cite as reasons when asked, and on how much more is part of their reasons.”¹⁵ Raz presents cases that he categorizes as “intentional actions for a reason,” but which are performed “semi-automatically,” “swiftly,” “opaquely,” etc. These cases seem not to sit well with the constructivist account of practical reasons, where properly reasons-driven action presupposes a demanding mental activity, in which, among other things, one’s proposed actions are examined concerning their fittingness and coherence with the agent’s principle-structured identity. Can there, according to this picture, ever occur cases “in which the reasons we act for are not the reasons which figure in our deliberations” (*ibid.*, p. 231)?

There can’t, *pace* Raz. Also with respect to the cases that Raz presents, it remains the case that, when it comes to the reasons for which agents act, deliberative endorsement and the mobilization of one’s principles are necessary features. In order to defend this position, I need to introduce another suggestion that will assign a role to practical principles that is more foundational than even many constructivists have considered necessary. I consider a wider formulation of the role that one’s self-conception and its principles play in the process of practical reasoning and action. The question then for the remainder of this paper is the extent to which a constructivist account can accommodate the instances of acts that Raz describes not merely as intentional actions, but as intentional actions done for reasons. I introduce two options that constructivism can put forward in order to accommodate Raz’s examples.

Korsgaard herself qualifies the role that deliberation plays in her account of acting for a reason: “In claiming that such [practical] identities are the sources of our reasons, I am not claiming that thoughts about your identity need to come

into your reasoning in any explicit way” (Korsgaard, 2009, p. 21). This qualification is noteworthy because it constitutes quite a significant accommodation of critics, such as Raz, who downplay the role of principles and deliberation in sufficiently accounting for reasons-based action. The problem is that Korsgaard further develops neither the qualification itself nor the implications that such a substantial concession has for her overall view. Given the dramatic and crucial function that actual deliberation in terms of principles has in the identity-constitution framework, one needs to elaborate more on her claim that even actions that do not invoke an agent’s normative self-conception “in any explicit way” nevertheless count as intentional actions done for reasons. There is a danger here of conceding too much.

And it is at this point that the first suggestion should be introduced: what I call the “indirect role of practical principles” is an alternative approach to render the overly intellectualist and rationalist picture of practical reasons and action more plausible. The apparent implausibility of this picture makes Raz’s examples of automatic, swift, and opaque actions attractive. In response to Raz (and Korsgaard’s weakening of her own claim) the first suggestion is that, even in the case of “semi-automatically, swiftly, and opaquely” chosen and performed actions, an agent’s practical identity *is* (though indirectly) a necessary component in bringing it about as an action for a reason. A fairly trivial semi-automatic action, such as heading to the fridge to get a drink because you’re thirsty, while not preceded by any explicit processes of practical deliberation, nevertheless remains *controlled* and checked by the confines set by one’s identity-constituting norms and principles.¹⁶

One way to look at this indirect function of agential identities is to observe how they not merely include positive commitments, aims, and projects, but are to a significant extent best conceived as negative principles that function like red lines circumscribing an agent’s options and actions. And these principles figure in the actions that Raz presents as supposed counterexamples to the constructivist requirement that all actions are principle-guided.¹⁷ This idea of assigning principles a much wider role in an agent’s practical life actually helps to see that constructivists and realists are not that far apart from each other. Since they always set the background limits on our actions, practical principles define and circumscribe “our appreciation of our situation in the world (an appreciation of ourselves as well as of our environment)” (Raz, 1999, p. 231). At the same time, there is no need for constructivists like Korsgaard to push any of these apparent concessions too far when it comes to the three types of cases of actions that Raz presents. Even if deliberative principles are not invoked explicitly and in the form of deploying them in self-aware practical deliberation, they remain a necessary background condition for these borderline cases of action to count as done for reasons. In the case of Raz’s examples of automatic actions (the fridge example, washing your hands before having a meal), *my* practical principles of etiquette ensure that I *don’t* sit down at the dinner table with dirty hands. The principles *are* in play and co-constitute the reason for washing my hands. Furthermore, in the case of opaque actions, such as the case of Raz’s agent who

decides to accept a job offer seemingly intuitively,¹⁸ deliberative principles, weighing the pros and cons of the considerations involved, co-inform the decision made. Lastly, and this example will occupy us the most in the remainder of our discussion, the swift action of rescuing a toddler from stepping into the busy road is circumscribed by moral and morally neutral normative principles that are part and parcel of my perception of the situation. The option of not taking the rescuing action is checked by principles that stand ready to jump in should my reflexive response to the situation make me hesitate to take the action in question. Again, this is probably the most controversial case and I will return to it in a couple of paragraphs.

Still, there remain problems in connection with Raz's examples for the constructivist, even if the reader accepts my argument from the indirect role of practical principles. Raz might rightly insist that there is a difference between deliberative principles passively standing ready in counterfactual scenarios, on the one hand, and actual deliberation in accordance and from these principles, on the other.¹⁹ This problem is also highlighted by the Korsgaardian slogan "action is self-constitution" (Korsgaard, 2009, p. 25). The self-constituting function is not merely an accidental feature of actions, but an essential one. Unless an act done for an end constitutes the person who performs it into its author, it is not an action. How then can actions constitute the agent who performs them when so many actions are done semi-automatically, swiftly, and opaquely? When it comes to the identity-constituting function of actions, one surely can't rest content with the merely indirect role of practical principles. Okay, Raz might grant, you have shown that even swift and opaque actions, in order to count as reasons-based, must somehow incorporate an appeal to an agent's practical principles—namely, in an indirect and passive mode. But constructivists (and especially constitutivists) want more: they claim that choice and action *from* those principles is what constitutes the individual agent into the unique individual that he or she is. And this seems to require some kind of active deliberation. Moreover, in order for behaviour to count as actions done for reasons, the constructivist must show that these actions fully satisfy this demanding standard and not merely incorporate deliberative principles in an indirect and passive manner.

My second suggestion is supposed to explain why Raz's borderline cases deserve to be labelled as actions done for reasons only when *at some point* they actively engage the agent's deliberative stance. However, for me to do so requires a bit of controversial action-theoretical work. The controversial bit of the suggestion consists in the possibility of an incorporation of reasons-based actions into one's practical identity after the outward manifestation of the action has already occurred. What is, for example, a reflexive action at one point (swiftly rescuing the toddler) *becomes* a full-blooded, reasons-based action only at a different point in time and only once the agent employs deliberative principles in the action's conceptualization. By allowing this alternative view of how reasons-based actions come about, I can preserve the constructivist claim that it is a defining feature of such action that it *does* depend on active practical deliberation in terms of principles.

I am well aware that this approach to self-constituting action seems to violate the view that reasons, similar to causes, always precede action. In order to render the proposal more plausible, let us take a cue from Hegel's remarks on the nature of action.²⁰ Hegel claims that an action is not necessarily completed at the time a specific empirical event that was initiated by an agent ends. An agent's intentions fully reveal themselves to others, but also to the agent himself or herself, only upon completion of an action, and this is a temporally extended process that reaches beyond the time span in which the empirical manifestation of the action takes place. Robert Pippin summarizes Hegel's view in the following way: "The deed and the reception and reaction to it are considered a *constitutive element* of the deed, of what fixes ultimately *what was done* and *what turned out to be* a subject's intention" (Pippin, 2008, p. 152). In a similar vein, my second suggestion is that an action's self-constituting power is not sufficiently accounted for when one ignores how it reverberates, so to speak, into the agent's continuing incorporation of it into his or her normative self-conception. In its self-constituting mode, the action's job is not done at a particular point in time, but continues to develop its significance in the course of practical deliberation and reflection—with oneself and in community with other agents. And with respect to the examples that Raz introduced, this suggests that it is not necessary for an "action that is done for a reason" to be performed in the deliberative and reflective mode at the point in time when it manifests itself in the world initially. What is necessary, however, is that this mode be activated at some point, with respect to the action in question.

Take again Raz's swift action of reaching out to the toddler who is about to step into a busy road (Raz, 1999, pp. 232-233). Raz has plausibly pointed out that "you do not stop to ask yourself what you should do. You react to the situation as you see it" (*ibid.*, p. 233). As already emphasized, he crucially adds that the agent in question does act for reasons, even if that agent does not have them in front of him or her at the time of initiating the action. Raz focuses on the forward-looking role of practical deliberation and concludes, as he does in the other cases mentioned: practical deliberation does not necessarily need to precede instances of actions for reasons. Firstly, he reminds us that the action in question is both intentional and done for reasons ("[reasons] which are available to them [the swiftly-acting agents] if asked at the time" (*ibid.*)). Secondly, the action is *not* preceded by any form of conscious deliberation and self-constituting activities. Hence, actions for reasons do not require deliberation and principles.

I would like to suggest a different account of what is going on in the toddler case. It is not the case that Raz's account of the example is incorrect, but it's simply not the whole story needed to make sense of the role of reasons. According to my conception of what acting for reasons amounts to, practical deliberation is very much a necessary post- (rather than pre-) condition of intelligible and reasons-based action, worth that demanding label. As emphasized above, acting for a reason is often considered to be a phenomenon incorporating a specific temporal structure, in which reasons are conceptualized analogously to "causes."

And since causes precede the effects that they bring about, it is only natural that Raz and others think that the examples of swift and semi-automatic actions are sufficient evidence for the idea that the activity of practical deliberation can be conceptually separated from counting someone as an agent who acts for a reason. If, however, the Hegel-inspired account suggests that the action of stopping the toddler is not, in all relevant senses, completed before the rescuer practically deliberates about it and thereby brings the reason for it into existence, I end up with a view about practical reasons that is not exclusively modelled within the causalist paradigm.

Let me clarify. I certainly do not want to endorse some implausible story about backward causation here (and neither does Hegel)—I am not arguing that the reason for rescuing the toddler, once it emerges in the process of continued self-constituting action, now suddenly becomes the psychologically efficacious cause that brought the rescue about. Rather, the picture is the following: action, according to the alternative view that differs from Raz's, in order to be action done for a reason *is* after all always connected with practical reasoning and deliberation that employ principles of choice and action. However—and this is the controversial move—this picture works only when, firstly, actions are regarded as having a temporally extended structure and, secondly, reasons for them can present themselves in a mode that contradicts the idea that practical reasons always work analogously to causes, especially concerning their temporal ordering.

I think, then, that the story about action qua self-constitution should be read in the following way. Engagement in practical deliberation, choice, and action does not mean that agents always engage in complex thought processes *before* they engage in actions that constitute them into their authors; that would indeed be an implausible and empirically unsustainable claim, and Raz rightly highlights cases of actions that, at the moment of performance, are done without being engaged in any processes of practical deliberation. However, these deliberative processes have to occur at some point in order for the actions in question to satisfy their self-constituting function and, importantly, to count as done for a reason. The rescuer who continues to fail to reflectively employ any principles (and, hence, generate reasons) that had led him or her to reach out and stop the child wouldn't count as constituting himself or herself successfully. And, contrary to Raz, that agent would not count as having performed an action based on reasons compared to an agent who does choose and act under a specific normative self-conception that incorporates that action. That this choice and genesis of reasons take place in the form of what appears to be a retroactive process need not be considered a problem once one gives up on an exclusively causalist framework. Reasons for action are not necessarily reasons *before* action (though very often they are). The presence and emergence of reasons for action continue to be dependent on first-personal and principle-structured deliberation. And this is true of all actions that are done for practical reasons.

There remain many issues. The first one leads us back to the question of how reasons are connected with the phenomenon of normative motivation. After all,

earlier in this section, I highlighted that the constructivist paradigm, due to its internalist commitments, presents a fairly convincing account of why practical reasons do not leave agents cold. The argument from how our normative self-conceptions are implicated in the processes of practical deliberation and choice was the main part of this account. It is, however, also with respect to the issue of practical reasons' motivational efficacy that the Hegel-inspired view about the temporal nature of practical reasons seems to run into a significant problem. After all, I certainly do not want that view to commit itself to the implausible claim that the motivational aspect of reasons can be created retroactively and *after* the action had been taken. The causalist intuition seems strongest at this point: the processes of practical deliberation that result in the respective reasons must be completed before the action and must manifest themselves in a resolve to engage in the action in question. In a sense, then, the Hegelian view intended as a charitable addition to the constructivist approach seems to go too far when it claims that engaging in practical deliberation is indeed *always* a precondition for reasons to emerge. What about the toddler case, for example? How can my version of internalism make sense of this and of the other cases that Raz presents as supposed counterexamples to the view that practical reasons are inherently connected to processes of deliberation? If it is not a reason that motivates the rescuer to perform his or her action, what is it?

The solution to that puzzle insists that even the motivating function of practical reasons is conditional on agents engaging in practical deliberation about what they do. The person stepping into the busy road to rescue the toddler is not moved and motivated to do so by a normative *reason*; rather, it is an immediate inclination to perform the rescuing act that does the motivating job in Raz's cases. The inclination must be incorporable into a practical reason, though. The proposed amendment to the internalist account can stick to its guns by insisting that, also with respect to the motivational dimensions of the person rescuing the toddler, it is incorrect to refer to a supposed motivating reason being in play at the moment the rescue takes place. For the motivating inclination to turn into a motivating reason, it is again necessary that a process of practical deliberation take place. Even with respect to normative motivation, this deliberation might take place after the rescuing act was performed. That bullet I am willing to bite. I can do so because inclinations take over the role of psychological efficacious causes in Raz's swift, opaque, etc. actions. Contrary to what Raz says, rescuing the toddler is not, at least not at all times of the action's genesis, an action done for a reason; and it is not, at least not throughout, an action motivated by the appreciation of some reason in favour of it. Also, motivating reasons must be the result of the deliberative deployment of one's identity-constituting principles that allow agents to endorse the inclinations with which they are confronted, and which result in the action in question.²¹

One last objection that I want to briefly address will probably emerge in response to the claim that, in cases of swift and opaque actions, no reasons, neither normative nor motivating, are present, unless principle-guided deliberation has been engaged in at some point. What about the accountability and responsibility prac-

tices of praising and blaming that agents engage in? After all, fellow agents seem to be justified when they praise the person rescuing the toddler for his or her action, aren't they? And, the objection continues, this shows that agents do, contrary to what I have been claiming in this section, consider the rescuer to act for a reason after all—maybe a subconsciously appreciated one, but one that is *there* in Raz's realist mode. And this intuition concerning holding agents responsible and praising them seems to remain stable even if one fully acknowledges the Razian assumption that no practical deliberation whatsoever has preceded the act of rescuing the child. Lastly, and more problematic for my suggestion, most of us would continue to praise the agent even if he or she were never to reflectively generate the reasons in the retroactive deliberation that the Hegelian version of internalism insists upon.

This objection from accountability-practices and attitudes can be sharpened when considering a tragic version of the toddler case. The rescuer gets hit by a bus and dies in the course of successfully pushing aside the toddler. Obviously, in this case, the deliberative engagement that lets the reason for the rescuing action emerge must remain absent, since the rescuer simply did not get the chance to reflectively endorse his or her rescuer-identity and, hence, the inclination that triggered the rescue fails to get transformed into a reflectively endorsed reason. According to my suggestion, the rescuer was directly and unreflectively acting on the basis of inclinations only; no reasons had yet entered the picture. But this claim seems implausible, given the third-personal evaluation of the case. When others praise the rescuer they presume that his or her agency was as fully engaged in the action as it is in the standard cases of action in which agents have the time and the cognitive resources to deliberate about what to do (and have been doing) and generate their reasons for action in a fully conscious manner (and without requiring the controversial Hegelian addition of retroactive endorsement).

I don't think that the counterintuitive impression that comes with the tragic version of the toddler case suffices to undermine the claim that the rescuer's action is indeed constituting an instance of "unfinished business" in so far as it lacks reasons. The version of internalism defended here leaves enough room for our accountability practices, even in cases of agents who are not able to engage in the task of retroactive reasons-genesis. (Interesting complications emerge in cases where agents fail to engage in that task for other reasons. Forrest Gump might be a case in which, while not completely absent, reflective endorsement is simplified and single-layered.)

In response to this objection from accountability, I should recall the first suggestion presented in response to Raz's critique of the importance of active deliberation. According to that suggestion, the rescuer *is* accountable for his or her praiseworthy action when that agent's agency-constituting principles at least passively and indirectly formed the inclinations and incentives that were the origins of his or her action. So, when we hold the rescuer accountable, at the very least we praise these normative practices and, more importantly, the kind

of practical agent that these principles constitute—not the bare inclinations. Still—and this must be repeated and insisted upon—we cannot hold the rescuer accountable on the grounds that he or she rescued the child on the basis of a proper normative and motivating *reason*. That the unfortunate rescuer fails to actively endorse the action-triggering inclination (that he or she forcefully experiences but in a non-reason-like manner) remains a certain deficit also with respect to the extent to which one can praise that agent.

Yet another strategy out of the current problem, which cannot be pursued here, is to reconsider the notions of incentive and inclination. It might turn out that their principle-structured identities not only always at least passively constrain what incentives and inclinations agents act upon, but also shape what they perceive as incentives and how they experience their inclinations to begin with. If that were the case, then the rescuer in the tragic version of the thought experiment turns out to be worth other's praise because his or her incentives and inclinations had been of a certain, normatively loaded, perceptual quality. Again, however, this quality must ultimately be the expression of the agent being of a certain kind—namely, one who actively endorsed good- and brave-making practical principles at some point. This explanation allows practical principles to take on yet another role in addition to their reasons-generating function—namely, in terms of principle-guided Aristotelian habituation. Obviously, these are rich issues that have to be further developed at another occasion.

6. CONCLUDING REMARKS

This paper's starting point was the idea to reconcile two conceptions of practical reasons that are widely discussed and considered to disagree on many of the fundamental issues in ethical theory and the theory of action. The ultimate outcome of this attempt is probably something entirely different—namely, an account of practical reasons that both Razian realists, on the one hand, and neo-Kantian constructivists, on the other, will be very critical of. The resulting account draws from both paradigms and combines the diverse elements with the aim of avoiding certain inconsistencies and problems that the original two strands of theories were struggling with.

I presented two main lines of argument to illustrate this alternative conception of practical reasons: the placebo analogy and a deliberation-focused account of how normative and even motivating reasons are generated. The placebo analogy's main claim was that agents have to (deliberately and deliberatively) uphold an inherently paradoxical conception of the relationship between themselves and their reasons. The apparent paradox emerges upon examining the presuppositions of the processes of efficacious practical deliberation and action. Reasons exist and they are *there*. However, they exist only because agents with their normative self-conceptions bring them about in processes of practical deliberation (I presumed that Korsgaard has presented a strong case for this throughout my argument). However, the same agents can engage in these processes in a way that is efficacious enough for them to successfully take up their roles as

practical agents only when part of their normative self-conception comes with the (principle-structured) stance that *negates* this exclusively constructivist, agency-dependent, character of reasons. As agents, we are forced to maintain this stance on practical grounds, exactly in order to coherently uphold anything before ourselves. We are *practically* justified in not rendering the nature of the relationship between ourselves and our reasons fully transparent.

The second claim was introduced to deal with cases of intentional actions that seem to involve agents who do not at all engage in these principle-governed processes of identity-constituting deliberation, which my view considers to be a necessary condition for reasons to be part of an action. In response to Raz's examples of swift, opaque, and automatic actions, I have insisted that reasons are not yet present, neither in their normative nor in their motivating varieties, because practical deliberation had not been engaged in. I have presented a noncausalist and Hegel-inspired alternative account of reasons-based action to show that reasons for a particular action do not always and necessarily precede the action that they co-constitute. If that account works, then it can successfully support the view that all practical reasons are the result of quite demanding deliberative processes in which an agent mobilizes the normative resources that he or she sees him- or herself committed to. This claim is certainly controversial regarding the normative features of our reasons, but has been especially difficult to defend in the case of reasons' role as motivating features of our agency. The hope is that the two suggestions taken together provide some support for this alternative conception of practical reasons and that it can be further developed into a view that overcomes a set of related controversies in contemporary ethical theory.

ACKNOWLEDGMENTS

I thank the audiences at the 2016 Eastern APA, at Ohio University, and at The University of Kentucky for their critical and valuable comments. Later versions profited a lot from the comments by two anonymous referees for this journal and from a colloquium discussion with Herlinde Pauer-Studer, Veli Mitova, and Alexandra Couto at the University of Vienna. Special thanks to Fabienne Peter for extensive written comments. Research for this paper was funded by the ERC Advanced Grant “Distortions of Normativity.”

NOTES

- ¹ My discussion of the principles-reasons connection focuses on two passages in Raz (1999, pp. 225-228; 2012, pp. 6-7). For Raz’s work on the nature of reasons more generally, see Raz (1999) and Raz (2012, Parts I and II). Regarding Korsgaard, my discussion singles out various passages throughout Korsgaard (2009).
- ² In this paper I put aside the so-called “paradox of self-constitution,” which Korsgaard acknowledges in her own work, and which has generated a lively debate in the secondary literature. In a nutshell, the paradox points out that in order to perform “self-constituting actions” (worth that demanding label), at the moment of action one has to already be a well-unified agent (i.e., someone with a principle-structured practical identity). But how can that be, if in order to *become* such an agent, one has to perform exactly these actions *first*? Whether Korsgaard’s response, that self-constitution must be regarded not as a state but as a dynamic process, is successful remains controversial (Korsgaard, 2009, pp. 41-44; Hanisch, 2013, pp. 103-107).
- ³ Korsgaard clarifies the complex interplay between incentives and inclinations, and their role in accounting for the phenomenon of a practical reason, by pointing out that one has to extend the analysis of choice in a temporal dimension: “Before the decision you seem to be asking whether the inclination provides a reason, while, after you make it, it seems to be the incentive that has provided it after all” (Korsgaard, 2009, p. 124).
- ⁴ The enormous complexity of Raz’s conception of a reason for action becomes noticeable at the following point, which should at least be mentioned. I am not sure how much hinges on the passage, but I take it that it supports the interpretation of Raz suggested in the text. Raz introduces the notion of an “explanatory reason” (in contrast to what one might call a “proper reason”) and defines it as “the agents’ belief that there is a reason for their action, or a specific belief of theirs in something which they take to be a reason can also be said to be a reason.” And the crucial passage reads: “An explanatory reason may exist even when there is no reason for the agents to do what they did” (Raz, 1999, p. 23 n.). Part of the point of the argument in the text is to come to terms with this distinction that Raz introduces here. “Somebody has a reason” is a notion that comes with a deep ambiguity.
- ⁵ I am indebted to a referee for this journal for highlighting that my initial presentation of Korsgaard and Raz as “causalists” about reasons was too strong. It evoked the incorrect impression that their causalism consists in an almost natural scientific understanding of reasons *qua* causes. The text now clarifies that the feature of their views that strikes me as problematic is one feature of the causal *analogy* in particular—namely, the temporal priority of reasons over actions. Both Raz and Korsgaard endorse this feature. In Raz’s case, this becomes visible when he discusses the cases of actions (also discussed in this essay) that are done for reasons that the agent is not aware of at the moment of action. These opaque reasons are still there and precede the action. See his discussion of “The Guiding Problem” (Raz, 1999, pp. 230-238). Korsgaard’s commitment to the temporal priority of reasons is even more explicit when she endorses and further develops Kant’s famous notion of the (practically free) will manifesting itself as “a causality of freedom” (Korsgaard, 2009, pp. 104-108). For a thorough reconstruc-

tion as well as critique of Korsgaard's central notion of "causality of the will," see Tubert (2011). On the post-Davidsonian debates regarding causalism more generally (and its lasting impact on how moral philosophers think about reasons for action), see D'Oro and Sandis (2013).

⁶ Recently, Korsgaard has endorsed a variety of this appeal to nature, noticeable when she says things like: "At the basis of every desire or inclination, no matter how articulately we can defend it, is a basic suitability-to-us that is a matter of nature and not of reason" (Korsgaard, 2009, p. 122).

⁷ The empirical side of how placebos actually work is not our main concern here. There are interesting studies that suggest that, under certain circumstances that still need to be investigated more conclusively, placebos do *not* cease to work even if the patient claims to have (fully) acknowledged the facts concerning the drug's complete inefficaciousness. The argument in the text presupposes an "idealized" placebo effect (Moerman 2002).

⁸ Sharon Street's (2008) "formalist" (as opposed to Korsgaard's "substantive") constructivism raises a number of distinct challenges for the placebo analogy. It is impossible to do justice to Street's distinct and complex theory, especially since the focus of this paper is Korsgaard's and Raz's positions. There is one brief discussion in particular, however, that hints at an important issue that distinguishes Street's constructivism from the placebo analogy. She says at one point that "evolutionary considerations, among others, show that to the extent that normative experience attributes any property of 'counting in favor of' to objects as they are in themselves, utterly independent of us and our attitudes, *that experience is in error*" (Street, 2008, pp. 240-241; my emphasis). In the footnote immediately accompanying this claim, Street continues to inquire in a bit more detail into this question of the extent to which our actual normative experience *does* attribute such an agency-independent property of "counting in favor of." Her conclusion: "This matter requires further discussion, however, and if the case can be made that normative experience attributes the property of 'counting in favor of' to objects as they are in and of themselves, utterly independent of us and our attitudes, then I embrace an error theory about the content of that experience" (Street, 2008, p. 241 n.). The placebo analogy both agrees and disagrees with the error-theoretical conclusion that Street suggests. True, qua metaethical theory, much speaks in favour of Street's evolutionary argument against realism (in the same sense in which medical, scientific, and pharmaceutical research speaks in favour of declaring the placebo's efficaciousness to be an error). And it is, therefore, also plausible to claim that our normative experience (in attributing realist properties to the objects of our practical deliberation) would to that extent commit an (ontological) error (in the same sense in which an agent who feels better after having taken the placebo pill is to that extent in an erroneous state). However—and this is the genuine contribution of the placebo analogy spelled out in this article—since successful agency is a thoroughly practical (as opposed to metaethical) affair, it is a misnomer to call the "realist illusion" that we must uphold in order to be robust agents an "error." Again, this aspect of Street's argument is embedded in a rich and complex theory of the nature of normative reasons. Clarifying the relationship between my hybrid view and her constructivism-all-the-way-down has to await another occasion and is certainly incomplete as presented here.

⁹ Elsewhere I (2016) investigate the question of whether appeals to inescapability are a promising avenue for vindicating unconditional normativity more thoroughly. I am indebted to a commentator for urging me to clarify the extent to which the placebo analogy results in a metaethical position that resembles Blackburn's (1993, pp. 166-181) quasi-realism, moral fictionalism (Joyce, 2001; Kalderon, 2005), and Gibbard's (1990; 2003) expressivism. The main differences between these theories in metaethics and my view are two. First, my practical realism is "practical through and through" and does not directly address the metaethical concerns that motivate Blackburn's and the fictionalists' accounts. Second, my account is concerned not with the foundations of morality but with the more general phenomenon of normative reasons for action, not necessarily moral ones.

- ¹⁰I cannot here discuss Raz's (1999, pp. 218-246) limited endorsement (but otherwise rather critical stance) toward particularism à la Dancy and its relationship to Raz's conception of incommensurability.
- ¹¹I (2013) argue for this part of the argument in more detail elsewhere.
- ¹²The seeds of a similar account regarding the first-personal trouble that threatens to ensue once agents thoroughly incorporate into their practical self-conceptions the Kantian/Korsgaardian story about reasons and values are also present in a paper by Fitzpatrick (2005).
- ¹³Part of my argument will remind the reader of David Enoch's (2007, pp. 35-41; 2011) recent defence of robust realism about practical reasons. It is impossible to fully spell out the relationship between the placebo analogy and Enoch's phenomenological claim that taking ourselves seriously as practical deliberators inescapably comes with a commitment to normative truths and facts (in terms of agency-independent reasons). The first part of my argument clearly resembles Enoch's strategy: first-personal deliberation must at least partly understand itself as aiming at realistically conceived items. However, the placebo analogy highlights that the first part of the argument does not necessarily entail realism. Rather, the "deliberative indispensability" that Enoch rightly highlights entails at most a disjunction (consisting of at least two disjuncts): normative realism is true *or* the placebo story is. The argument in this text of course tries to defend the latter but cannot present any additional arguments undermining the plausibility of the former. That would require an exhaustive discussion of Enoch's argument.
- ¹⁴Humean and Kantian accounts of practical reasons differ dramatically in many aspects, but are nevertheless rightly categorized as internalist. For a good recent statement of this categorization, see Scanlon (2014, pp. 1-15).
- ¹⁵Raz (1999, p. 230). A bit of context might be necessary, my qualification in the text notwithstanding. Raz tries to refute the view, argued for by Dancy, that there is a gap between an agent's reasons, on the one hand, and all the evaluative considerations applying to the agent's situation, on the other. Raising the guiding problem supports Dancy: "Only those evaluative considerations which can serve a guiding function are reasons for action" (Raz, 1999, p. 230). Raz then goes on to argue that this gap does not in fact exist: all considerations pertaining to the evaluation of actions are, at the same time, reasons for/against these actions.
- ¹⁶On the issue of conscious control, see the constitutivism defended by J. David Velleman (1996, pp. 714-726).
- ¹⁷Elsewhere I (2013a) introduce the paradoxical-sounding notion of "negative goals," in order to clarify this confining function of our practical identities.
- ¹⁸Raz's category of opaque actions is actually a rather complicated matter, because it combines the idea of principle-guided deliberation at time *t*, on the one hand, and unprincipled decisions (though based on what had happened at *t*) at *t'*. Raz's example presents an agent who deliberates about taking up a job offer, who allows that deliberative process to sit a couple of days, and who then makes the decision. The decision is an opaque one because at the point of actually deciding the matter, "I do not deliberate. Nor am I able, once the decision is taken, to adduce any reasons to justify it that I did not have before. I have no new arguments. I just made up my mind, apparently on the basis of the old arguments" (Raz, 1999, p. 233). My response in the text is that, in such opaque cases, the deliberative principles employed earlier in the decision process also co-constitute the reason for which the job offer is ultimately accepted.
- ¹⁹In discussing the fridge example, Raz gets very close to acknowledging the line of argument that I am going to develop in the text. He says: "They [automatic actions] are all done for reasons, but these have been instilled in me and I need not reflect on them or re-endorse them. The reason-guided character of the actions is manifest in the fact that I monitor them, and will abort them if the situation changes, or is revealed to have changed" (Raz, 1999, p. 232). It is the requirement of "reflection" and "re-endorsement" of reasons (that result from *active* practical deliberation as opposed to mere monitoring) that I want to defend as a necessary feature of automatic actions that, at the same time, count as "reason-guided."

²⁰The two main sources in Hegel that contemporary analytic philosophers single out as the inspiration for the argument I rehearse in the text are Hegel's remarks on the concept of action and agency in the *Phenomenology of Spirit* (especially the second half of Chapter Five) and *Elements of the Philosophy of Right* (Part Two: Morality). See Hegel (1977, p. 211-262) and Hegel (1991, pp. 133-142). In addition to the work by Pippin cited in the text, see the agenda-setting monograph by Quante (2004) and the anthology by Laitinen and Sandis (2010) that incorporates essays that contrast the Hegelian view of action with the Davidsonian, empiricist and causalist, contenders in discussions regarding reasons for action.

²¹I am, once again, indebted to a referee for reminding me that the constructivist account of practical reasons for swift and other, more reflective but still rather pedestrian, actions has a tendency to "over-intellectualize" and "over-aggrandize" the role of identity-defining principles and self-conscious reflection regarding one's actions. This is certainly a prominent concern, and common sense, as well as many philosophical perspectives (including Raz's), seems to favour the view that not fully rational creatures and human agents can also perfectly well act for reasons without engaging in demanding deliberative activities. One strategy in response to this objection is to allow agency and reasons-based action to come in degrees. I pursue this more nuanced strategy elsewhere (Hanisch, 2013, pp. 124-136). Still, this alternative constructivist argument also ultimately insists that a sophisticated level of transparent self-reflection *is* necessary for a high degree of successful identity-constitution and reasons-based action. As the concluding reflections on responsibility practices in the text try to show, the seemingly over-intellectualizing tendencies of my account should become less implausible when we consider the case of an agent who performs Raz's swift action and *never* engages in reflectively incorporating this action into his or her normative personality. The Hegelian alternative helps here because we certainly do not want to require the rescuer to engage in complex identity-related deliberation *before* he or she steps into the busy street and reaches for the toddler. This variety of the one-thought-too-many objection emerges only when we adhere to the framework that requires reasons for the rescuing action to be present before the action begins. My claim in this section has been that we do not end up with a problematic "one thought too many" if the conception of reasons-based action defended here requires the rescuer to engage in practical deliberation *after* the reflexive components of his or her (praiseworthy) action are completed. (My remark concerning Forrest Gump highlights my awareness that much more needs to be said about the issue of degrees of praiseworthiness of the spontaneous rescuer.)

REFERENCES

- Anscombe, G. E. M., *Intention*, Cambridge MA, Harvard University Press, 2000.
- Blackburn, Simon, *Essays in Quasi-Realism*, New York NY, Oxford University Press, 1993.
- D'Oro, Giuseppina and Constantine Sandis (eds.), *Reasons and Causes: Causalism and Anti-Causalism in the Philosophy of Action*, London, Palgrave Macmillan, 2013.
- Enoch, David, "An Outline of an Argument for Robust Metanormative Realism", in Shafer-Landau, Russ (ed.), *Oxford Studies in Metaethics (Volume 2)*, Oxford, Clarendon Press, 2007, p. 21-50.
- , *Taking Morality Seriously. A Defense of Robust Realism*, New York NY, Oxford University Press, 2011.
- Fitzpatrick, William J., "The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity", *Ethics*, Vol. 115, No. 4, 2005, p. 651-691.
- Gibbard, Allan, *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, New York NY, Oxford University Press, 1990.
- , *Thinking How to Live*, Cambridge MA, Harvard University Press, 2003.
- Hanisch, Christoph, *Why the Law Matters to You*, Boston MA, DeGruyter, 2013.
- , "Negative Goals and Identity: Revisiting Sen's Critique of *Homo Economicus*", *Rationality, Markets, and Morals*, Vol. 4, 2013a, p. 157-172.
- , "Self-Constitution and Inescapability: A Diagnosis", *Philosophia*, Vol. 44, No. 4, 2016, p. 1145-1164.
- Hegel, G. W. F., *Phenomenology of Spirit. Translated by A.V. Miller*, New York NY, Oxford University Press, 1977.
- , *Elements of the Philosophy of Right. Translated by H.B. Nisbet*, New York NY, Cambridge University Press, 1991.
- Joyce, Richard, *The Myth of Morality*, New York NY, Cambridge University Press, 2001.
- Kalderon, Mark Eli, *Moral Fictionalism*, New York NY, Oxford University Press, 2005.
- Korsgaard, Christine M., *The Constitution of Agency. Essays on Practical Reason and Moral Psychology*, New York NY, Oxford University Press, 2008.
- , *Self-Constitution: Agency, Identity, and Integrity*, New York NY, Oxford University Press, 2009.
- Laitinen, Arto and Constantine Sandis (eds.), *Hegel on Action*, London, Palgrave Macmillan, 2010.
- McDowell, John, *Mind and World: With a New Introduction*, Cambridge MA, Harvard University Press, 1996.

Moerman, Daniel E., *Meaning, Medicine, and the 'Placebo Effect'*, New York NY, Cambridge University Press, 2002.

O'Neill, Onora, *Constructions of Reason. Explorations of Kant's Practical Philosophy*, New York NY, Cambridge University Press, 1989.

—, *Towards Justice and Virtue. A Constructive Account of Practical Reasoning*, New York NY, Cambridge University Press, 1996.

—, *Acting on Principle. An Essay on Kantian Ethics*, New York NY, Cambridge University Press, 2013.

Parfit, Derek, *On What Matters*, New York NY, Oxford University Press, 2011.

Pippin, Robert B., *Hegel's Practical Philosophy – Rational Agency as Ethical Life*, New York NY, Cambridge University Press, 2008.

Quante, Michael, *Hegel's Concept of Action*, New York NY, Cambridge University Press, 2004.

Raz, Joseph, *Engaging Reason*, New York NY, Oxford University Press, 1999.

—, *From Normativity to Responsibility*, New York NY, Oxford University Press, 2012.

Regan, Donald H., "The Value of Rational Nature", *Ethics*, Vol. 112, No. 2, 2002, p. 267-91.

Scanlon, T. M., "Reasons: A Puzzling Duality?", in Wallace, R. Jay, Philip Pettit, Samuel Scheffler and Michael Smith (eds.), *Reason and Value. Themes from the Moral Philosophy of Joseph Raz*, New York NY, Oxford University Press, 2004, p. 231-246.

Scanlon, T. M., *Being Realistic About Reasons*, New York NY, Oxford University Press, 2014.

Street, Sharon, "Constructivism about Reasons", in Shafer-Landau, Russ (ed.), *Oxford Studies in Metaethics (Volume 3)*, Oxford, Clarendon Press, 2008, p. 207-245.

—, "Mind-Independence without the Mystery: Why Quasi-Realists Can't Have It Both Ways", in Shafer-Landau, Russ (ed.), *Oxford Studies in Metaethics (Volume 6)*, Oxford, Clarendon Press, 2011, p. 1-32.

Tubert, Ariela, "Korsgaard's Constitutive Arguments and the Principles of Practical Reason", *The Philosophical Quarterly*, Vol. 61, No. 243, 2011, p. 343-362.

Velleman, J. David, "The Possibility of Practical Reason", *Ethics*, Vol. 106, No. 4, 1996, p. 694-726.

A STRINGENT BUT CRITICAL ACTUALIST SUBJECTIVISM ABOUT WELL-BEING

STÉPHANE LEMAIRE

MAÎTRE DE CONFÉRENCES EN METAÉTHIQUE ET PHILOSOPHIE DE L'ESPRIT, UNIVERSITÉ DE RENNES I

ABSTRACT:

Subjectivists about well-being claim that an object is good for someone if and only if this individual holds a certain type of pro-attitude toward this object. In this paper, I focus on the dispute among subjectivists that opposes those who think that the relevant pro-attitudes are *actual* to those who think that they are counterfactual under some idealized conditions. My main claim is that subjectivism should be stringently actualist, though our actual pro-attitudes may be criticized from an intrinsic perspective. To defend this claim, I first present three desiderata that a subjectivist theory of well-being should fulfil. Two of these desiderata result from the fact that a subjectivist theory of well-being should not be implicitly paternalist, while the other is that it should be able to play a normative role. I then show that several actualist theories that introduce light forms of idealization or other conditions that have a similar aim, fail to satisfy at least one of the antipaternalist desiderata. This gives some legitimacy to a very stringent version of actualism. I then describe three features of the kind of stringent actualism that I want to defend, which will explain the ability of what is good for someone to play its normative role. Finally, I show how these features allow us to deal with the classic objection that the objects of actual “defective attitudes” cannot be good for the holder of these pro-attitudes.

RÉSUMÉ :

Les subjectivistes à propos du bien-être soutiennent qu'un objet est bon pour un individu si et seulement si cet individu possède un certain type d'attitude positive à l'égard de cet objet. Dans cet article, je me concentre sur le débat à l'intérieur du camp subjectiviste qui oppose ceux qui pensent que les attitudes positives pertinentes sont actuelles à ceux qui pensent qu'elles sont contrefactuelles sous une condition d'idéalisation. Ma thèse principale est que le subjectivisme devrait être rigoureusement actualiste bien que nos attitudes positives actuelles puissent être critiquées d'un point de vue qui leur est interne. Afin de défendre cette position, je présente d'abord trois *desiderata* qu'une théorie subjectiviste du bien-être devrait satisfaire. Deux de ces *desiderata* résultent du fait qu'une théorie subjectiviste du bien-être ne devrait pas être implicitement paternaliste, alors que le troisième résulte de ce qu'elle devrait pouvoir jouer un rôle normatif. Je montre ensuite que plusieurs théories actualistes qui introduisent des formes modestes d'idéalisation ou d'autres conditions ayant la même visée échouent à satisfaire au moins un des *desiderata* anti-paternalistes. Une version très rigoureuse de l'actualisme subjectiviste s'en trouve ainsi légitimée. Je décris ensuite les trois aspects de l'actualisme rigoureux que je souhaite défendre et qui permettront d'expliquer dans quelle mesure ce qui est bon pour un individu peut jouer un rôle normatif. Enfin, je montre comment ces aspects nous permettent de répondre à l'objection classique selon laquelle les objets d'« attitudes déficientes » actuelles ne peuvent être bons pour ceux qui ont ces attitudes.

1. INTRODUCTION

According to subjectivism about well-being, something is good for an individual if and only if this individual has a certain type of pro-attitude *A* toward this object under certain conditions *C*. Within this broad framework, there are many debates concerning the precise analysis that subjectivism should offer. For instance, subjectivists disagree about what type of pro-attitude is relevant. Some have suggested that it should be an attitude of desiring (Sidgwick 1907, Murphy 1999; Heathwood 2005) or of valuing; others, that it should be a second-order desire (Lewis 1989, Railton 1986a); still others, that it should be a belief about what is good for one (Dorsey 2012). However, I will not be concerned with this debate in the present paper. Rather, I will focus on the dispute between those subjectivists who think that the relevant pro-attitudes are *actual* pro-attitudes and those who think that they are *counterfactual* under some idealized conditions. More precisely, the aim of the paper is to defend a new version of actualist subjectivism that is stringently actualist while explaining how this stringent actualism can fulfil the normative role that what is good for someone plays. I will try to achieve this aim while staying as neutral as possible on all the other debates among friends of subjectivism.

The plan of the paper is the following: Section Two starts with the presentation and defence of three desiderata that a subjectivist theory of what is good for someone should fulfil. In a nutshell, the desiderata derive from two considerations that push in opposing directions. On the one hand, an important drive toward subjectivist approaches to well-being is to have a theory such that someone's well-being is not detached from what matters to him or her. On the other hand, what is good for someone often plays a normative role: it is a tool to criticize our actual pro-attitudes. Putting these desiderata to work, I show in Section Three that some versions of actualism are not sufficiently actualist. This will be my plea for a very stringent form of actualism. However, the traditional problem with stringent actualism has been its inability to fulfil its normative role. In order to overcome this problem, I will present three features of the stringent actualism that I want to defend. One of these features opposes my stringent actualism to all other forms of stringent actualism. I will call this actualism Stringent Critical Actualism. In Section Five, I will finally show how the three features of Stringent Critical Actualism allow it to deal with the classic "defective" pro-attitude objection that the objects of some of our actual pro-attitudes may not be good for us. Section Six concludes the paper.

2. THREE DESIDERATA OVER SUBJECTIVISM

As mentioned above, the debate that opposes actualist to idealized approaches of subjectivism has appealed to two considerations. The first consideration is normative. We often take what is good for someone to depart from what that person desires, believes is good for him or her, etc. For instance, one may entertain desires whose fulfilment will be disappointing if not catastrophic. Hence, it

seems that a subjectivist needs to be able to assess the pro-attitudes that we have *de facto*. Historically, this has been a strong argument against actualist versions of subjectivism.¹

A second consideration has received much attention: it is the idea that what is good for someone must matter to him or her. Railton's often-cited remark that "it would be an intolerably alienated conception of someone's good to imagine that it might fail in any such way to engage him"² seems to capture this idea. Let us call it the No-Alienation requirement. This requirement is certainly a main rationale that drives us toward subjectivist versus perfectionist or objective theories of well-being. The defect of the latter theories seems to be that by their very definition of well-being, they imply that some objects can be good for an individual while he or she is in fact indifferent to these objects. And this seems—at least to many—to pave the way to unacceptable forms of paternalism as it might be considered better to provide people with objects to which they are completely indifferent rather than with objects that matter to them for their good. Therefore, if subjectivist approaches to well-being are seen as superior because they satisfy a No-Alienation requirement, then this requirement becomes at least a *prima facie* constraint to which a subjectivist theory of well-being must conform.³ Notice, moreover, that the No-Alienation requirement as used by Railton or Dorsey, for instance, is also an argument against idealized forms of subjectivism. Indeed, such idealized versions of subjectivism have the same defect that traditional perfectionist and objectivist views have, since insofar as they also end up saying that what is good for someone may be indifferent to him or her as he or she actually is. And this underlines that the proper interpretation of the No-Alienation requirement has to be rather strict: it is a necessary condition for an object to be part of the good of a person that this object would matter to this person if he or she were considering it.

While these first two requirements are important, I want to suggest that we need to add what I will call a No-Rejection requirement. The idea is simple and is somehow symmetrical with the No-Alienation requirement: at least in some cases, it seems unacceptably paternalistic to act toward someone as if something that mattered to him or her counted for nothing. Why? Intuitively, the problem is that if something matters to someone, then it must count as part of that person's well-being even if it may be true that this thing will in the end be disappointing. Consider, for instance, a young man who strongly wants to become a philosopher and who is enthusiastic about everything that concerns philosophy. However, suppose that he lacks a clear understanding of what a philosophical career consists in, and that, if he were able to consider it properly, he would not desire it anymore. Finally, suppose that he is presently unable to consider this career properly. If you tried to explain to him what the career consists in, he would just say that you were wrong, that it cannot be right, and would get upset and annoyed by your unwelcome explanations. I want to claim that in such a case his enthusiasm must be taken as implying that the philosophical career is good for him. Why? Because in doing so, we have the most natural ground to explain why it is wrong to act toward such people without taking this enthusi-

asm into account, which is that a certain part of what is good for them is neglected. In a nutshell, because some actions may be seen as unacceptably paternalistic in virtue of the fact that they neglect what matters to someone, we should accept a No-Rejection requirement on well-being, which precisely says that what matters to someone is good for this person.

Before going further, let me be clear about the status of the two last constraints that I have put forward. My point is not to argue directly that we must accept a certain theory of well-being because some forms of paternalism are wrong. My point is merely to argue that a *subjectivist* theory of well-being should *allow* for the rejection of certain actions as wrong and paternalistic because they do not take into account what matters to people. In other words, all things being equal, we should discard subjectivist theories of well-being that would be rejected by anti-paternalists on the grounds that these theories do not tie tightly enough what is good for someone to what matters to him or her. And I argue that the two constraints considered above should be accepted precisely as expressing this theoretical orientation.

I acknowledge that this point has little dialectical force against objectivists or perfectionists, who can respond that they are just trying to describe what well-being consists in, and that the legitimacy of paternalism is a different point. But my aim here is not to argue directly against these approaches to well-being. My point is merely that *if* one is tempted to reject these views because they are seen as potentially allowing⁴ some form of unacceptable paternalism, and if one is thereby tempted by a subjectivist approach, then the two antipaternalist requirements considered above express this orientation. In other words, I am only trying to offer arguments that should be seen as relevant to thinkers who are already within the subjectivist camp and who believe that certain antipaternalistic considerations may be relevant.

This being said, notice that the normative requirement seems at first sight to pull exactly in the opposite direction insofar as it emphasizes that a theory of well-being must offer the tools for an assessment of what presently matters to us. However, it does not follow that the three requirements are incompatible. In fact, the aim of this paper is precisely to argue that if one is attracted by a subjectivist approach to well-being, then one should accept all these requirements and therefore a stringently actualist version of subjectivism insofar as it is the only one which satisfies them all. If successful, such a program will offer a theory of well-being that could be a shared ground for discussions between paternalists and antipaternalists insofar as this theory does not exclude antipaternalists by adopting a theory of well-being that is unacceptable to them.

I therefore conclude that a subjectivist theory of what is good for someone must satisfy both a normative requirement and two antipaternalist requirements: the No-Alienation requirement and the No-Rejection requirement. The aim of this paper is to offer an actualism that satisfies all three desiderata. To justify this actualism, I will highlight, in the next section, that several actualist views of

what is good for someone are mistaken because they are not stringent enough. These objections will pave the way for the Stringent Actualism that I wish to offer.

3. NOT SUFFICIENTLY STRINGENT ACTUALISMS

In this section, I show that we have to reject certain forms of actualism because they do not satisfy at least one of the two antipaternalist requirements. As a consequence, I will also show that ideal pro-attitude forms of subjectivisms must be rejected.

3.1 Semi-actualist views

Let us first consider what I will call semi-actualist versions of subjectivism. I refer here to the approaches suggested by Sobel (2009) and Dorsey (2017) that have been constructed with the intent to satisfy the No-Alienation requirement. Their key point is to introduce a distinction between two roles that idealization plays in subjectivist theories of well-being. Dorsey presents the distinction as one between the two building blocks that constitute a subjectivist theory of what is good for someone. The first building block is what he calls the “Good-Value Link.” He states it as follows: “for any object, event, state, etc., O and agent x, O is good for x only if O is valued, under condition c, by x” (Dorsey 2017, p. 200). He argues that it is with regard to this link that alienation must be avoided. Therefore, he defends an actual version of this link to the effect that, for any object, event, state, etc., O and agent x, O is good for x only if O is valued, under *actual* condition c. However, a subjectivist account of well-being is incomplete until one has offered an explanation of what it is to value something, hence the need for a second building block that explains what it is to value something. According to Dorsey, it is at this stage that idealization should be introduced. In his view, to value an object O cannot be reduced to the actual possession of a certain type of pro-attitude toward O. Rather, an individual x values an object O only if x would have a certain type of pro-attitude toward O *under some idealization conditions*. Why? As Sobel has suggested, taking desire as the relevant pro-attitude, the crucial point is that our actual desires may result from a misappreciation of their object. Hence, “desires that do not involve [a] perfect forecast are, in a sense, not actually *for the option as it is* but rather for the option as it is falsely imagined to be” (Sobel 2009, p. 345. Italics are mine). On the contrary, “the desire is truly for [O] when the desire is sustained or created in light of complete and accurate information about what [O] would be like” (*ibid.*, p. 346). Along similar lines, Dorsey claims that “a necessary condition for x to value [O] is that x would take the relevant pro-attitude toward [O] given full consideration of the relevant range of imaginative conditions, i.e., ways [O] might be” (Dorsey 2017, p. 209). In addition, he claims that the pro-attitudes must also belong to the pro-attitudes of x were these pro-attitudes rendered minimally coherent.⁵ The main reason he offers for this second idealizing condition is that “when someone maintains incoherent pro-attitudes, it can be unclear, ultimately, what he values” (*ibid.*, p. 204).

I contend that the distinction between the Good-Value Link and the Theory of Valuing Link is largely artificial when we consider the two antipaternalist requirements. In order to make the argument clearer, I will focus on the view that values should be explained in terms of desires. I believe that the argument offered here would apply *mutatis mutandis* to other analyses of values in terms of other pro-attitudes. To justify my contention, I will return to the young man who desires with great enthusiasm to become a philosopher. We have supposed that his desire is not for the career as it truly is. Nevertheless, if one tried to correct his understanding of what is in fact involved in embracing a philosophical career, one would feel that he somehow does not want to be informed. He would simply resist any change in his beliefs about the nature of a philosophical career. However, his present refusal may at some point vanish and he would then realize that he has no desire to become a philosopher. Able to fully consider a different career, he would in fact desire to become a political activist, a desire which he does not hold presently.

Now, let us see what semi-actualists would say about this case. First, insofar as the present desire would not persist if the young man were accurately informed, being a philosopher has no value for the young man. Second, full consideration would lead him to desire to be a political activist, so Sobel would claim that this is what has value for him. Dorsey, however, defends a slightly different view: on his view, for an object to be good for someone, it is necessary but not sufficient that this person desire this object while accurately informed. It is also necessary that the person presently desire this same object. It follows that, on Dorsey's view, it is not good for the young man to embrace a philosophical career in virtue of the desire that he would have if accurately informed insofar as the young man is not also presently desiring this career.

How do these characterizations of what is good for the young man fare with regard to our two antipaternalist requirements? As we will see, neither satisfies both requirements. Start with Sobel's view: first, he claims that the political career is good for the young man. But to accept that is just to violate the No-Alienation requirement insofar as the young man does not feel presently engaged by a political career. This is why Dorsey's response is preferable. Turning now to the philosophical career, both Sobel and Dorsey claim that this career is not good for the young man insofar as he would not still desire it if he were accurately informed. Hence, both violate the No-Rejection requirement.

In addition, I would like to suggest that the whole idea of a full consideration of objects in order to reveal what is good for someone is misguided. Our desires, our pro-attitudes, and even our beliefs about what is good for us rely always on a partial understanding of their objects. In fact, a large part of our life consists precisely in moving toward or turning around the objects of our pro-attitudes and thereby discovering their nature from different perspectives, their consequences, and whether or not they continue to be of interest to us. Consider the simple case of wine tasting. Surely, the whole good of it is not to have the wines we like and will continue to like if we were experts. That could be boring. Part

of the good that we derive from our experience of wines is progressively to acquire some knowledge about what we truly like, to be surprised about wines we thought we do not like (but in fact do like), and even to discover, interestingly, that we dislike some wines, or do not like them any more—or at least not as much as we used to like them, etc. In other words, the enterprise of acquiring knowledge about the objects of our desires through discoveries, surprises, and disappointments is part of what is good. And an important part of that is also to become less attracted to what seduced us at first. This seems to be a truism about the development of a culture in art. Hence, it is not true that what would be good for us would just be to have objects that we would still desire when fully aware of their exact nature. This would remove all the good we take in our first encounters with objects, in experiencing them, discovering them, and leaving them, in orienting and reorienting ourselves toward and away from what seems to us from moment to moment as having worth.

But this is not all. I want to claim that even if we put aside the good in acquiring knowledge and in dynamic experience, it is a mistake to assess the good of an object with reference to a unique, ideally informed point of view. To see this, consider the good of a certain friendship, a loved person, or a specific occupation. It seems inappropriate to assess how good they are for someone on the basis of a unique, ideally informed perspective. Consider the beginning of a love story or a new friendship. Certainly part of what is good lies in our ignorance of the other. To assess this good as if we knew the other perfectly would seem like assessing how the love story or the friendship would be a long time after it had started. Even this comparison is mistaken because, then, time will have modified these relationships. In other words, I would like to contend that it is a mistake to hold that one can assess how good an object is for a person at a given time through the assessment of what would be the pro-attitude of this person if he or she were ideally informed about this object. The only way we can assess how good something is for someone is through the good it offers at various moments, given the specific history that has related that person to the object. Therefore, the very idea of a complete and vivid consideration is not what we need to clarify what is good for us. What we need is our actual pro-attitudes toward an object as they may evolve through time while our knowledge improves—but also while what is relevant about these objects evolves. As I will argue in Section Four, this is not to say that there is no place for the remark that some objects are disappointing. For the time being, my point is merely that in designing a theory of value, we should not start by discarding badly informed pro-attitudes, and this, for two reasons. First, it seems very harsh to claim that the first music or wines or people that we loved were not good for us insofar as our experience has led us away from them. Second, the very idea of a complete and vivid representation of something is not what is relevant when we want to know why an object is or may be good for us in the present or in the future: our knowledge is always partial and, even when it becomes less so, it seems more proper to say that the object is good to us now in a different way than to say that we have reached the true way in which it is good to us.

Therefore, I conclude that Sobel's and Dorsey's semi-actualist views face important objections. First, they both violate the No-Rejection requirement. Second, they both mistakenly consider that an object cannot be good for someone if this person does not have the relevant pro-attitude toward it while accurately knowing this object. Finally, and in addition, Sobel's view violates the No-Alienation requirement.

3.2 Impure actualism

Let us now turn to more rigorous versions of actualism that do not involve any kind of idealization. In this subsection, I would like to show that even some such rigorous forms of actualism still do not satisfy the No-Rejection requirement. This will be done by focusing on the actualism defended by Heathwood (2005).

As an actualist version of subjectivism in terms of desires, this view claims that an object *O* is good for an individual *A* if and only if *A* actually desires *O*. This does not settle the supplementary question as to whether a desired object benefits the person who has this desire when the desire is fulfilled or when the person had the unfulfilled desire. Heathwood himself defends the claim that an object benefits the individual who desires it at the moment at which the desired object is obtained, a view that may be called after Dorsey (2013) the "time-of-object" view. On the contrary, an opponent to this view could defend a "time-of-desire" view (again, after Dorsey 2013), which claims that the object desired benefits the individual who desires it at the time at which he or she desires it and not at the time at which it is obtained. As it will appear, we won't be concerned with this debate. Suffice it to say that in both cases it is claimed that it is impossible to benefit a desire if the desire is not fulfilled.

Finally, Heathwood introduces an additional condition for an object *O* to be good for someone, which he calls *Concurrence*: "the desired object *O* is desired *at the same time* that it is obtained" (adapted from Heathwood 2005, p. 490).⁶ For instance, "if I desire fame today but get it tomorrow when I no longer want it," fame is not good for me (*ibid.*, p. 490). Notice that Concurrence is compatible with both the time-of-desire and the time-of-object views. It just adds a condition for the fulfilment of a desire to benefit its possessor: that the possessor still has the desire when it is fulfilled.

All that being said, we can appreciate an interesting consequence of Concurrence. To do so, we can notice first that, at the very time at which a desire is satisfied, its object is obviously more vividly experienced. At least, if one still desires the object while one experiences it, then it seems that the object is not disappointing. Thus, if it is required, for an object to be good, that one still desire that object while the desire is being fulfilled, then it follows that if an object is good for someone, then it cannot be disappointing for him or her. It is therefore a strength of Heathwood's view that it avoids the objection that the fulfilment of a desire may not be good for its possessor. Moreover, this objection is avoided without making use of any idealizing condition.

Unfortunately, the introduction of Concurrence, which itself presupposes that an object cannot be good for someone if it is not fulfilled, is unacceptable. This is because the object of a desire can benefit the possessor of the desire even if the desire is never realized. As it is well known, it is beneficial to someone who possess a desire to become quite sure that his or her desire is going to be satisfied in the future *even if this desire is not satisfied in the future*. Symmetrically, doubts about the satisfaction of a desire that concerns my child after my death diminishes my well-being. So it seems that it is an error to claim that the object of a desire brings no good if the desire is not fulfilled.

Now, there are two objections that one might raise. The first is to suggest that the experiences produced by our anticipations are just pleasures and displeasures. However, this response cannot be accepted within the subjectivist framework. We can understand these anticipations only in terms of the desire that we have for a certain object. Within this framework, it seems that the most plausible option is to slightly enlarge the desire-satisfaction view and to claim that someone benefits from a certain object that is desired either when the desire is fulfilled or when the possessor of the desire anticipates this fulfilment.

A second objection is to contend that this argument succeeds only for subjectivists who adopt a Knowledge condition that the fulfilment of a desire benefit its possessor only to the extent that he or she knows that it is fulfilled.⁷ But, as the argument goes, not all subjectivists adopt this Knowledge condition: many of them hold, on the contrary, that someone's well-being is improved not by knowing that his or her desire is satisfied, but merely by having the desire satisfied, whether or not this fact is known. In response, I would like to show that an argument for the same conclusion can be offered even if the Knowledge condition is rejected. Let us start with a typical intuition put forward by those who reject the Knowledge condition. Consider Peter, who desires to be esteemed by his relatives and who believes falsely that he is. According to those who reject the Knowledge condition, his situation is bad insofar as his desire is not fulfilled. But now, suppose that, although his desire is not already fulfilled, it is close: for some reason, his relatives are collectively more positive about him. It seems that we have to say that his situation has improved. Maybe the situation is not as good as it would be if his desire were satisfied, but it is nevertheless much better than it would be if the desire were far from being satisfied.

Now, it may be said that we are making a category mistake: Peter's situation is no better in absolute terms; rather, it is merely instrumentally better. But I am not sure about this objection. Consider situations in which someone got very close to the fulfilment of his or her desire, but didn't fulfil it. We think that such a situation is worse than the mere absence of fulfilment. But certainly this situation is not instrumentally worse. Therefore, it seems that the situations in which one does not fulfil a desire are not all equal in terms of noninstrumental goodness-for. Therefore, even subjectivists who are against the introduction of a Knowledge condition should accept that someone's well-being can improve or

decline depending on that individual's situation relative to the fulfilment of a certain desire. This is true of well-being independently of whether the desire is presently fulfilled and independently of whether the desire is fulfilled in the end.

The first problem that we have raised against Concurrence and its implications leads to a second. Recall the young man who enthusiastically desires now to be a philosopher, although he will completely lose this desire when he is about to fulfil it. It follows from Concurrence that the object of his desire is not good for him, and that this is a violation of the No-Rejection requirement. Once again, we can see that there is something mistaken here by considering how we should act toward this young man. If we follow the subjectivists who adopt Concurrence, then the philosophy career is not good for the young man. Hence, we lack a reason to take this desire into account and this may allow actions that we might consider as unacceptably paternalistic. Certainly, one might argue that the desire could be taken into account on other grounds. But what other grounds? The grounds of respect will be too broad, and if we say that we should respect his desire, are we not saying that we should take into account what is good for him? For it is hard to think that respect tells us to take into account what is not good for someone. In other words, the simplest and most plausible way to say why we should take into account his desire is precisely to acknowledge that if a philosophical career presently matters to the young man, then it is presently good for the young man.

I believe that the lesson of the present assessment of Heathwood's view is the following. Even if a condition such as Concurrence allows a stringent actualist to deny that the object of a disappointing desire is good for someone who desires it, the stringent actualist must reject this condition. This is because this stringent actualism fails to satisfy the No-Rejection requirement, and this in part because it fails to fully understand how we benefit from having desires. Does this leave open the possibility for an actualist to retain a condition in parallel with Concurrence, while allowing that one might benefit from the goodness of an object in virtue of a desire toward this object that will not be fulfilled or will be disappointing? Yes, it does. One might propose to replace Concurrence by Concurrence*, which would read as follows: "The object of a desire is good for the possessor of the desire only if the possessor possesses the desire at the moment at which he or she benefits from this object." Such a clause seems to me plausible. However, Concurrence* does not have the supposedly theoretical advantage that made the initial attraction of Concurrence since the former does not imply that the objects of disappointing desires are never good for the possessors of these desires. I conclude from the assessment of Heathwood's view that an actualist must avoid any idealization strategy, whether direct or indirect, and that we have at least good preliminary reasons to look for a very stringent and straightforward version of actualism.

4. STRINGENT ACTUALISM AND NORMATIVITY

From the previous section, it has appeared that subjectivism must be stringently actualist and must therefore merely assert the following:

An object, state of affairs, event, O is good for X if and only if X has actually a pro-attitude of type A toward O.

All this is good. However, one might wonder whether such a stringent actualism can satisfy the normative requirement. Is it faithful to the idea that the determination of what is good for someone, of his or her values, is a tool to *improve* his or her life? Or, as Railton puts it, is such a stringent actualism capable “of capturing important elements of the critical and self-critical character of value judgments” (Railton 1986b, p. 11)?

My aim in the next section is to show that a positive answer can be offered. In this section, I will present three features that allow a stringent actualism to play this normative role. The second and third features are shared by all actualisms, but I am going to argue that the first feature distinguishes the kind of stringent actualism that I want to defend from others. I will call this view Stringent Critical Actualism. Along the way, this presentation will allow me to show how this actualism accords with a set of intuitions about how we may help others and how people should think in order to improve their lives. The next section will put in practice the normative dimension of Stringent Critical Actualism and show how it deals with “defective” actual pro-attitudes that have been seen as counterexamples to stringent forms of actualism.

The first feature that I want to present has been rejected by most stringent actualists. They have argued that, if what is good is analyzed in terms of desires, then at least the underived or intrinsic desires cannot be assessed from an internal perspective. In particular, Murphy claims, “What does lie beyond criticism in terms of well-being, on the simple desire-fulfilment theory, is our deep, actual desires—whatever those happen to be” (Murphy 1999, p. 267).⁸ Hubin (1996) makes a similar claim. Although he is offering a theory not of well-being but of reasons in terms of desires, he is an actualist in claiming that an analysis in terms of counterfactual desires should be rejected in favour of an analysis in terms of actual desires. But what is relevant for us is that, in defending this reduction of reasons to actual desires, a key point in his view is that intrinsic motivations cannot be criticized.

With regard to theories of well-being with which I am concerned here, the thesis that intrinsic motivations cannot be criticized turns out to be false—for two reasons. First, if what is good for someone is at least in part to fulfil that person’s desires, then it is surely better to have more desires fulfilled. And from this, it follows that sets of desires can be assessed as more or less easy to fulfil, or as more or less coherent, in a way that allows the fulfilment of all or at least all of those that are of great importance. Hence, even if a desire taken in isolation

cannot be assessed, it can be assessed given the abilities of its owner to fulfil it and given the full set of desires that its owner possesses.⁹ The argument extends easily if the relevant pro-attitudes are second-order desires, attitudes of valuing, beliefs concerning what is good for one, etc.

Second, my defence of stringent actualism in the previous section allows us to make a further criticism of our actual pro-attitudes or desires, and it is here that my view definitely departs from Heathwood's and Murphy's views. I have argued that even if a desire for an object will appear disappointing or is badly informed, the object is good for the person. Hence, I contend consistently both that a badly informed desire makes its object good, but that this lack of information is criticizable. It is preferable to have better-informed desires insofar as the lack of information may lead to disappointment. In other words, I concede to Sobel and Dorsey that our actual present pro-attitudes or desires may rely on a lack of consideration of their objects, even though I do not conclude, with them, that this is sufficient for their objects to lack value. So I depart from both kinds of actualism: I claim that the objects of our present actual desires are good for us, even if the desires or pro-attitudes will be disappointing. But the ground of this claim is not that our intrinsic desires cannot be disappointing or cannot be intrinsically criticized.¹⁰ On the contrary, I claim that we can criticize desires or pro-attitudes either because they lack coherence or because they rely on an insufficient or wrong understanding of their objects.

Notice, in addition, that this view offers an original response to the objection of Enoch (2005) that any idealization of our actual pro-attitudes to define well-being relies implicitly on objective goods. Following Sobel, I claim against Enoch that some idealizations or criticisms that are intrinsic to desires or pro-attitudes make sense. However, as I have now argued at length, I do not follow Sobel in claiming that the objects of these idealized desires or pro-attitudes are the constituents of our well-being. To this extent, I agree with Enoch that if subjectivism is defensible, it must be stringently actualist. However, as already said, this is not because our actual pro-attitudes or desires have authority by default.

So much for the first feature of Stringent Critical Actualism. Let us then turn to the second feature that will allow it to fulfil the normative requirement. This is the simple fact that actualism does not preclude the assessment of what is good for someone and thereby of his or her pro-attitudes with respect to a variety of extrinsic norms and scales, by which I mean norms or scales that are not relevant to assess someone's well-being. In particular, one can assess whether those things that are good for someone are in accordance with moral demands, whether these goods have intellectual or aesthetic worth, etc. Here my point is not to make a claim about these extrinsic norms. Nor is it to decide whether they are true norms or just scales that have no normative force. Nor is it even to explain which of these norms have precedence over others. Rather, my point is merely to insist that Stringent Critical Actualism is fully compatible with the possibility of assessing someone's goods not only from an internal but also from numerous external perspectives.¹¹

The third and final important feature on which Stringent Critical Actualism may rely to satisfy the normative requirement derives from the temporality of pro-attitudes: insofar as pro-attitudes change over time, what is good for someone changes accordingly. Admittedly, this is an obvious point that applies to all subjectivist accounts. If it is held that O is good for x if and only if x has a certain type of pro-attitude toward O under certain relevant conditions, then a temporal change in the actual or counterfactual pro-attitudes of x is a temporal change in what is good for x. However, it is not clear that all subjectivists have adopted this temporal approach. The point is especially relevant with regard to idealized versions of subjectivism because the idealizing element may pull toward a conception of individual values as temporally stable. Suppose you define what is good for someone as what this person would desire having full information, if this desire belonged to a maximally coherent set of desires. In this configuration, it is tempting to suggest that what is good for someone at a certain time is also good at any other time. After all, if there is a set of desires that is somehow optimally rational, it seems that it is impossible to improve on it. Therefore, it seems¹² that what is good for you at a given time is what is good for you at any other time.¹³ In any case, my point here is merely to emphasize that in contrast with idealized subjectivism, which may be tempted—even if not committed—to claim that our idealized set of pro-attitudes would be stable across time, actualist subjectivisms and thus Stringent Critical Actualism are committed to accepting that what is good for someone changes as his or her relevant pro-attitudes change. Hence, what is good for someone during his or her lifetime must not be restricted to what is good for that person at a certain given time. What is good for someone in his or her lifetime is the integration of what is good for that person at all the different times of his or her life, whichever view one may take on how these temporally different goods should be integrated.

These three features put us, I believe, in the best possible position to assess someone's well-being and how beneficial it would be to attempt to improve it. First, it is possible to assess someone's set of present actual pro-attitudes as providing less good than it could and to think of other sets of pro-attitudes that this same person could acquire and that would improve his or her well-being in the future. But second, insofar as someone's present actual pro-attitudes constitute what is good for him or her at this present moment, they cannot be set aside as worthless. Third and finally, since changing one's pro-attitude usually has a cost, this will also need to be taken into account. Thus, attempting to improve one's life is not a decision that involves leaving non-goods for goods—in which case, very costly but small improvements may be taken as rational. Rather, what the decision has to balance is the costs involved in changing my present actual set of pro-attitudes and the benefits of having a new set of pro-attitudes that will improve my well-being.

Similar remarks apply to the assessment of actions intended to contribute to someone else's well-being. Two questions will need to be answered. First, is it worthwhile to improve the set of pro-attitudes of this individual in view of the cost that the improvement itself may carry? Second, are we allowed to impose

this cost, or even to interact in order to improve the pro-attitudes of the individual under consideration? The same considerations will apply if we consider a moral improvement of an individual's set of pro-attitudes. Are the moral benefits really worthwhile? Are we allowed to impose some costs or to interfere to improve the morality of this set of pro-attitudes?

My aim is certainly not to answer the second question and its moral variation. These are important and further questions that go beyond our understanding of well-being. Nevertheless, I contend that Stringent Critical Actualism is the only subjectivist theory of well-being that offers a common ground to discuss the legitimacy or illegitimacy of paternalist actions and generally actions in favour of someone's well-being.

5. THE DEFECTIVE PRO-ATTITUDE OBJECTIONS

It is now time to see how Stringent Critical Actualism satisfies the normative requirement when confronted with objects of defective pro-attitudes. Actualist versions of subjectivism in terms of desires have been criticized as unable to deal with so-called defective desires whose objects we do not intuitively want to take as good for someone. Among these defective desires, we find ill-informed desires such as the desire to drink this glass that contains petrol or the desire for a career that is grounded in an insufficient knowledge of what this career consists in; there are pointless or "quirky"¹⁴ desires, such as the desire to turn on radios as soon as one sees them (Quinn 1993) or to count the blades of grass in the lawn (Rawls 1971); there is also the desire of the former drug addict to fall back into addiction, etc. Obviously, the objection extends *mutatis mutandis* to other versions of actualism expressed in terms of other pro-attitudes. So, how does Stringent Critical Actualism deal with so-called defective pro-attitudes?

A lack of space precludes a full discussion of the long list of defective pro-attitudes that may be relevant. I will not consider here desires that are considered as defective in virtue of extrinsic norms (such as moral norms), insofar as they are not relevant to the assessment of well-being. I do not want to rehearse a point here that most subjectivists have made. That being said, Stringent Critical Actualism can also rely on responses that other actualists have already offered in order to deal with some cases of defective desires that I will briefly consider first.¹⁵ Second, I will consider the cases in which Stringent Critical Actualism has an original answer, and especially the case of pro-attitudes that rely on insufficient information about their object.

To begin, let us consider the objection raised by quirky desires or pro-attitudes. There are two responses that stringent actualists may offer. The first is to underline that the relevant pro-attitude in the subjectivist analysis of what is good is not the pro-attitude of desiring, but either a cognitive attitude or a conative pro-attitude, such as intrinsically desiring or valuing. If it is the pro-attitude of believing that something is good for one, the objection collapses since we do not have incomprehensible beliefs, such as the belief that it is good for me to turn on all

the radios I come across. Similarly, the objection collapses if the pro-attitude is an intrinsic desire or an attitude of valuing. Indeed, if we consider our intrinsic desires, it seems that their objects are never seen as completely external to us. Similarly, if an object is valued, then the desire for this object cannot be an event completely detached from us.

Let us then turn to the case of the former drug addict. According to Stringent Critical Actualism, it is good for the former drug addict to have his or her drug because the drug is the object of, say, a desire. But it is often argued that it cannot be good for the drug addict to have his drug. My response is in two parts. First, I see no reason to deny that Stringent Critical Actualism is giving the proper response. I do not see why we should deny that satisfying the desire for the drug is good for the drug addict to some extent. Nevertheless, I admit that the fulfilment of this desire could be bad, all things considered for the drug addict. There is nothing problematic here insofar as this good may certainly be inferior to other goods for him or her that are incompatible with this addiction. Moreover, it is certainly better for the drug addict to have a future set of desires that excludes this desire, since fulfilling it will always be in conflict with his or her other desires. Hence, because Stringent Critical Actualism can assess a set of desires as offering someone a life that could be improved, it can make sense of the intuition that satisfying this desire will be bad for the drug addict on the whole and that it would be better for him or her not to have this desire at all. Nevertheless, admitting all this does not require us to deny that the drug is presently and actually good for the drug addict. The point is that he or she should merely reject this good in order to live a better life.

Finally, let us turn to cases of ill-informed pro-attitudes or desires. In the context of this paper, it is crucial that Stringent Critical Actualism deal properly with such cases since we have rejected as inappropriate views that have been designed precisely to deal with them. I believe that in order to manage them, we should distinguish two types of cases.

The first type of case is that where one holds a mistaken instrumental belief or a mistaken belief that a certain object specifies a certain more general one. This is the case where I want to drink this glass of liquid because I am thirsty, but ignorant that it contains petrol. Are stringent actualists forced to concede that drinking petrol is good? Certainly not. The Stringent Critical Actualist can claim, as other actualists have, that an object is good for someone only if this individual has an underived or intrinsic desire or pro-attitude toward this object.¹⁶ So, it is easy to deal with these cases insofar as the person in question has no underived desire to drink petrol, but rather an underived desire to drink something that will quench his or her thirst, and this excludes petrol.

This response is, however, insufficient when we consider the second type of case, the case that involves a deeply erroneous pro-attitude because it is far from obvious that such a desire is derived.¹⁷ Let's go back to the young man who desires to become a philosopher, although the desire relies on an erroneous,

partial, or abstract but well-entrenched understanding of this career, and who would be disappointed if he were to fulfil his pro-attitude. Is the object of this pro-attitude good for the young man? According to Stringent Critical Actualism, it is, and I think that this is the correct response. It may even be an important good that presently contributes heavily to the well-being of his life. He thinks a lot about it, is excited by the prospect when he discusses philosophical topics, and he takes pleasure in viewing himself as a philosopher in the future. Hence, I believe that insofar as being a philosopher is a good for him from which he presently benefits, this good may demand some attention from us. On the other hand, Stringent Critical Actualism can appeal to temporal considerations: if it is true that he will be disappointed by such a career and will thus not desire it anymore if he were to become a philosopher, then being a philosopher will not be good for him at some point in the future. Hence, the balance of his present and probable future goods may justify actions intended to discourage him, maybe by indirectly allowing him to discover on his own what the job that he presently values so much would truly be like, or by offering alternative possibilities whose value might be more lasting.

6. CONCLUSION

The stringent actualism that I have offered here relies on two points that have been insufficiently highlighted. The first is that the No-Alienation requirement is not sufficient to protect a theory of well-being from being implicitly paternalistic. We must add a No-Rejection requirement whose precise characterization should be the object of further work. In any case, when this requirement is taken into account, then it offers a strong rationale for adopting an actualist approach to subjectivism against any idealizing condition or any other condition (such as Concurrence) that goes in the same direction. The second point is that if the No-Rejection requirement is taken as a main rationale for being actualist, then we must concede that pro-attitudes, and especially desires, are intrinsically assessable: they may rely on a faulty representation of their object or they may be incoherent and lead thereby to a life endowed with less well-being. Given these possible assessments, a stringent actualism is not stuck with a view that claims the pro-attitudes cannot be intrinsically assessed. And, therefore, it can satisfy the normative requirement.

NOTES

¹ Sidgwick (1907); Brandt (1996); Railton (1986a and b).

² Railton (1986b, p. 9).

³ A similar point is emphasized by Dorsey (2017, p. 196).

⁴ I say “*potentially* allowing” because other considerations can lead a theory of right conduct to reject as wrong any action that would not take into account what matters to someone while denying that what matters is good for him or her. However, it is unclear what such considerations could be. Moreover, it is hard to believe that neglecting what matters would not be relevant to the assessment of the action as wrong.

⁵ In fact, Dorsey insists at several places that he does not want to commit himself to the idea that these two idealizing conditions are the only ones. Others may also be legitimate.

⁶ Heathwood (2005) also defends what has been termed, after Rabinowicz and Österberg (1996), a *satisfaction view* and not an *object view* within subjectivist approaches. This is a debate about the bearers of the property *good for*. On the object view, it is the object of our desires that is good for us. On the satisfaction view, what is good for us is to have desires satisfied. In this latter case, the relation of the property *good-for* is the state that conjoins a desire for an object and the object itself. In any case, this distinction is not relevant for us here. It is orthogonal to all the other distinctions made in this summary of Heathwood’s view.

⁷ In fact, Heathwood accepts a Knowledge condition. Thus, I would not need to consider this objection if he were my only target. However, insofar as my aim is to show that an actualist defending Concurrence is mistaken, I want to extend my critique to an actualist subjectivist who would adopt Concurrence but reject a Knowledge condition.

⁸ Heathwood (2005, p. 492) holds the same view.

⁹ In fairness to Heathwood, I must underline that he acknowledges this first kind of criticism, although he takes it to be external (Heathwood 2005, p. 494).

¹⁰ See, for instance, Murphy (1999, p. 268) for this claim.

¹¹ Certainly, it might be argued that some of these norms are relevant and thereby intrinsic to the assessment of well-being. However, I stick here to the subjectivist tradition, which denies the relevance of these norms as determining how well someone’s life is going. The aim of this paper here is not to defend subjectivism against alternative views such as perfectionism.

¹² At least, in most cases. In fact, desires may change due to aging or to hormonal factors in such a way that the most coherent and fully informed set of desires that could be derived from someone’s *present actual* desires would be different from the set of most coherent and fully informed desires that could be derived from the same individual’s *past actual* desires.

¹³ Smith (1994) has even argued that this set of desires would be the same for all humans.

¹⁴ For this term and a discussion of this specific class of defective desires, see Bruckner (2016).

¹⁵ See especially Heathwood (2005); Hubin (1996); Murphy (1999).

¹⁶ The strategy has been adopted by Hubin (1996), Murphy (1999), and Heathwood (2005).

¹⁷ Here I partly join Mendola (2009), who is dubious about the possibility for actualists to use the distinction between underived and derived desires. While I agree with this criticism in the kind of case we are now considering, it seems to me that in the first type of case, considered above, there is no difficulty in seeing them as derived.

REFERENCES

- Brandt, Richard B., *A Theory of the Good and the Right*, 2nd ed., Oxford, Clarendon Press, 1998.
- Bruckner, Donald W., “Quirky Desires and Well-Being”, *Journal of Ethics & Social Philosophy*, vol. 10, no. 2, 2016, p. 1-34.
- Dorsey, Dale, “Desire-Satisfaction and Welfare as Temporal”, *Ethical Theory and Moral Practice*, vol. 16, no 1, 2013, p. 151-171.
- , “Idealization and the Heart of Subjectivism”, *Noûs*, vol. 51, no. 1, 2017, p. 196–217.
- , “Subjectivism without Desire” *Philosophical Review*, vol. 121, no 3, 2012, p. 407-442.
- Enoch, David, “Why Idealize?”, *Ethics*, vol. 115, no. 4, 2005, p. 759-787.
- Fletcher, Guy, “A Fresh Start for the Objective-List Theory of Well-Being”, *Utilitas*, vol. 25, no. 2, 2013, p. 206-220.
- Heathwood, Chris, “The Problem of Defective Desires”, *Australasian Journal of Philosophy*, vol. 83, no. 4, 2005, p. 487-504.
- Lewis, David, “Dispositional Theories of Value”, *Proceedings of the Aristotelian Society, Supplementary Volumes*, vol. 63, 1989, p. 89-174.
- Mendola, Joseph, “Real Desires and Well-Being”, *Philosophical Issues*, vol. 19, no. 1, 2009, p. 148-165.
- Murphy, Mark C., “The Simple Desire-Fulfillment Theory”, *Noûs*, vol. 33, no. 2, 1999, p. 247-272.
- Quinn, Warren, “Putting Rationality in its Place”, in R. G. Frey and C. Morris (eds.), *Values, Welfare, and Morality*, Cambridge, Cambridge University Press, 1993, p. 26-50.
- Rabinowicz, Wlodek et Österberg, Jan, “Value Based on Preferences”, *Economics and Philosophy*, vol. 12, no. 1, 1996, p. 1-27.
- Railton, Peter, “Moral Realism”, *The Philosophical Review*, vol. 95, no. 2, 1986a, p. 163-207.
- , “Facts and Values”, *Philosophical Topics*, vol. 14, no. 2, 1986b, p. 5-31.
- Rawls, John, *A Theory of Justice*, Cambridge, Harvard University Press, 1971.
- Rosati, Connie S., “Persons, Perspectives, and Full Information Accounts of the Good”, *Ethics*, vol. 105, no. 2, 1995, p. 296-325.
- Sidgwick, Henri, *The Methods of Ethics*, 7th ed., Cambridge, Harvard University Press, 1907.
- Smith, Michael R., *The Moral Problem*, Hoboken, NJ., Blackwell Publishing, 1994.
- , “Internal Reasons”, *Philosophy and Phenomenological Research*, vol. 55, no. 1, 1995, p. 109-131.
- Sobel, David, “Subjectivism and Idealization”, *Ethics*, vol. 119, no. 2, 2009, p. 336-352.

LES ATTITUDES APPROPRIÉES *VERBATIM*

FABRICE TERONI

PROFESSEUR ASSOCIÉ, DÉPARTEMENT DE PHILOSOPHIE, UNIVERSITÉ DE GENÈVE

JULIEN DEONNA

PROFESSEUR ASSOCIÉ, DÉPARTEMENT DE PHILOSOPHIE, UNIVERSITÉ DE GENÈVE

RÉSUMÉ :

Selon l'analyse FA des concepts évaluatifs, notre conception d'un objet comme ayant une valeur donnée est la conception d'une certaine attitude évaluative appropriée à son endroit. Cet article examine deux défis que doit relever cette analyse. Le défi psychologique exige de l'analyse qu'elle fasse appel à des attitudes qui soient à même d'éclairer nos concepts évaluatifs, tout en ne présupposant pas la maîtrise de ces mêmes concepts. Le défi normatif réclame quant à lui que la compréhension du caractère approprié des attitudes pertinentes s'articule autour d'une forme de normativité intimement liée à la nature de ces attitudes, sans pour autant dépendre de la maîtrise des concepts évaluatifs analysés. Afin d'établir si l'analyse FA peut relever ces défis, nous commençons par clarifier la nature des attitudes auxquelles il convient de faire appel. Dans la première section, nous considérons des attitudes génériques comme les désirs et présentons des raisons de privilégier des attitudes plus spécifiques, à savoir les émotions. La suite de la discussion évalue la plausibilité d'une analyse FA souscrivant à l'une ou l'autre des approches contemporaines en théorie des émotions. La deuxième section souligne qu'il est difficile d'associer l'analyse FA aux théories en termes de *contenu* évaluatif et la troisième présente les atouts d'une théorie en termes d'*attitudes* évaluatives. Dans la dernière section, nous revenons au défi normatif et esquissons une manière de le relever.

ABSTRACT:

According to the FA analysis of evaluative concepts, our conception of an object as having a given value is the conception of a given evaluative attitude towards it as being appropriate. In this paper, we explore two challenges that this analysis has to take up. The *psychological* challenge requires that (i) the analysis appeal to attitudes that are apt to shed light on our evaluative concepts while (ii) not presupposing the mastery of these same concepts. The *normative* challenge requires that (i) we understand the appropriateness of the relevant attitudes in terms of a kind of normativity intimately related to these attitudes, again (ii) without this presupposing the mastery of the relevant evaluative concepts. To determine whether the FA analysis can meet these challenges, we start by clarifying the sort of attitudes the analysis should appeal to. In section 1, we consider generic attitudes like desires and give some reasons to favor more specific attitudes, namely emotions. The ensuing discussion assesses the prospects of an FA analysis in light of contemporary accounts of emotions. In section 2, we argue that the FA analysis should not be combined with accounts according to which emotions have evaluative *contents* and then offer, in section 3, some reasons for combining the FA with an account in terms of evaluative *attitudes*. In section 4, we show how this allows for solving the normative challenge.

Selon une approche courante, l'approbation et la désapprobation appropriées permettent de comprendre les valeurs – on parle alors d'analyse des valeurs sous l'angle des attitudes appropriées (en bref et pour coller à la terminologie anglaise, « analyse FA »).¹ Cette approche se décline en deux versions, l'une *ontologique*, l'autre *conceptuelle* (Tappolet, 2011). Selon la version ontologique, un objet possède une valeur donnée si et seulement si une certaine attitude à son endroit est appropriée. On dira par exemple qu'une action est bonne si et seulement s'il est approprié de l'approuver ou, autrement dit, si et seulement si elle mérite notre approbation. Selon la version conceptuelle, notre conception d'un objet comme ayant une valeur donnée est la conception d'une certaine attitude évaluative appropriée à son endroit. Concevoir bonne une action, c'est estimer qu'elle mérite notre approbation. Ces deux versions sont indépendantes l'une de l'autre et nous allons ici nous concentrer exclusivement sur la version conceptuelle. Or, si l'analyse FA des concepts évaluatifs est attrayante puisqu'elle promet de lever le mystère qui les entoure en les ancrant dans nos attitudes, elle doit néanmoins relever deux défis.

Premièrement, la promesse d'une analyse FA est d'élucider les concepts évaluatifs par les concepts d'attitudes appropriées. Cela contraint à éviter deux écueils qui constituent ce que nous appellerons le *défi psychologique*. D'abord, il ne faut pas que notre compréhension des attitudes en jeu dans l'analyse FA présuppose une maîtrise des concepts évaluatifs en question. Pour illustrer la difficulté, remarquez que si approuver un objet revient à juger qu'il est bon, alors les promesses de l'analyse FA ne peuvent être tenues. En effet, juger qu'un objet est bon présuppose la maîtrise du concept du bon. Il incombe donc au défenseur de la FA d'identifier des attitudes dont la compréhension est indépendante de la maîtrise de concepts évaluatifs. C'est là l'écueil bien connu de la *circULARITÉ* (Rabinowicz & Rønnow-Rasmussen, 2004).² Le second écueil est celui de l'*informativité* : les attitudes en jeu dans l'analyse FA ne doivent pas être trop éloignées des valeurs, sans quoi on ne verrait plus pourquoi la compréhension des valeurs présuppose la compréhension de ces attitudes. Imaginons par exemple que l'analyse ait recours à de purs *quale*, à savoir des états conscients qui ne représentent rien. Dans la mesure où ces états ne semblent pas être des formes d'approbation ou de désapprobation, pas plus d'ailleurs que des états auxquels s'applique la notion de mérite en jeu dans l'analyse FA, faire appel à notre compréhension des *quale* ne peut pas éclairer nos concepts évaluatifs de la manière envisagée par cette analyse.

Deuxièmement, une fois ce défi psychologique remporté, l'analyse doit relever un *défi normatif*. On bute ici également à deux écueils. D'abord, il ne faut pas que la compréhension du caractère approprié des attitudes pertinentes dépende de la maîtrise des concepts évaluatifs visés par l'analyse. Supposons que l'attitude d'approuver soit compréhensible indépendamment du concept de bon. Si notre concept d'approbation appropriée est celui d'approbation du bon, l'analyse FA demeure circulaire. C'est la déclinaison *normative* de l'écueil de la *circULARITÉ*. Ensuite, il ne faut pas non plus que la normativité à laquelle on a recours soit totalement étrangère à la nature des attitudes en jeu dans l'analyse. Suppo-

sons que le concept de raisons d'approuver soit celui d'un ensemble d'injonctions divines – ce qui mérite l'approbation est ce que Dieu nous enjoint d'approuver. À nouveau, on ne comprend alors pas l'intérêt d'un détour par les attitudes : nos concepts de bien et de mauvais peuvent être directement élucidés par notre concept d'injonctions divines. Cette version normative de l'écueil de l'*informativité* n'est que trop rarement reconnue (Dokic & Lemaire, 2015).

Afin d'évaluer si une analyse FA peut relever ces défis, la première étape consiste à clarifier la nature des attitudes couramment admises dans l'analyse. On distingue deux familles d'approches qui diffèrent par les concepts évaluatifs qu'elles ciblent et, corrélativement, les attitudes psychologiques auxquelles elles font appel. Les approches de la première famille ciblent les concepts de ce qui est bon et mauvais. Afin d'analyser ces concepts de valeurs dites *fines* (« thin »), elles ont recours à des attitudes *génériques* comme l'amour et la haine ou le désir et l'aversion. La seconde famille d'approches analyse les concepts de valeurs dites *épaisses* (« thick ») comme l'offensant, le honteux ou l'admirable en matière d'attitudes plus *spécifiques* comme la colère, la honte ou l'admiration.³

L'une de ces approches peut-elle élucider nos concepts évaluatifs en relevant les défis psychologique et normatif? Dans la première section, nous considérons le développement d'une analyse FA selon l'angle d'attitudes génériques et présentons des raisons de privilégier des attitudes plus spécifiques, à savoir les émotions. La suite de notre discussion évalue la plausibilité d'une analyse FA souscrivant à l'une ou l'autre des approches contemporaines en théorie des émotions. La section 2 souligne qu'il est difficile d'allier une analyse FA aux théories focalisant sur le *contenu* évaluatif, alors que la section 3 présente les atouts d'une théorie selon les *attitudes* évaluatives. Dans la section 4, nous revenons au défi normatif et esquissons une manière de le relever. Nous espérons ainsi démontrer que certains débats en méta-éthique profitent d'un examen approfondi de leurs soubassements psychologiques.

1. FA : ATTITUDES GÉNÉRIQUES OU SPÉCIFIQUES?

Faire appel à des attitudes génériques revient à appliquer l'analyse FA aux seuls concepts évaluatifs fins, afin d'opérer au sein du domaine évaluatif la distinction entre ce qui est bon et ce qui est mauvais. Mais comment comprendre les notions d'approbation et de désapprobation de façon générique si ce n'est en considérant les jugements évaluatifs dont nous avons vu qu'ils conduisent l'analyse dans une impasse? Au moins deux couples d'attitudes sont envisageables : *l'amour* et *la haine* (Brentano, 1889) d'un côté, et, de l'autre, *le désir* et *l'aversion* (Oddie, 2005). Ces attitudes ne semblent pas se ramener à des pensées évaluatives à propos de leurs objets et donc éviter l'écueil de la circularité. Dans cette section, après avoir dressé un bref portrait des approches selon les attitudes génériques, nous donnons des raisons de penser que l'analyse FA des concepts évaluatifs *fins et épais* doit faire appel aux émotions.

Selon une conception courante et plausible, l'amour et la haine sont des dispositions complexes à désirer et éprouver de façons spécifiques en fonction des circonstances dans lesquelles se trouve l'objet aimé ou haï (Broad, 1954; Deonna & Teroni, 2009; Naar, 2013). Si Sam aime Maria, il est disposé à désirer que ses projets se réalisent, à ressentir de la joie en sa présence, de la tristesse lorsqu'elle se trouve loin de lui, de la colère lorsqu'elle est offensée, etc. Or, quelle que soit la manière dont on développe cette idée, une analyse FA se focalisant sur l'amour et la haine repousse le problème plutôt qu'elle ne le résout. En effet, les concepts d'amour et de haine formés par le sujet dépendent de la façon dont ces dispositions se révèlent à lui. Dans la mesure où les dispositions psychologiques ne lui sont pas immédiatement accessibles, les concepts résultants sont ceux de ce qui explique ou cause les émotions et les désirs pertinents. Autrement dit, l'analyse FA des concepts évaluatifs fins selon l'amour et la haine présuppose la viabilité d'approches FA considérant les désirs ou les émotions : si ces dernières s'avéraient non circulaires, il serait envisageable de les appliquer aux dispositions que sont l'amour et la haine. C'est donc vers ces analyses qu'il faut tourner notre attention.

Quelles sont les chances de succès d'une analyse FA des concepts évaluatifs fins ayant recours au couple désir-aversion? L'idée sous-jacente est la suivante : les désirs sont la source principale de notre conception du bon dans la mesure où ils nous motivent à poursuivre certaines fins, la conception résultante du bon étant celle de *ce qui mérite d'être poursuivi*. C'est une option à première vue séduisante, car elle ne repose pas sur des attitudes explicitement évaluatives et évite ainsi l'écueil de la circularité. Cependant, son succès dépend entièrement de la notion de désir en jeu. A-t-on affaire à des désirs conçus comme des dispositions ou comme des occurrences? S'il s'agit de dispositions, on s'expose bien sûr aux soucis déjà entrevus à propos de l'amour et de la haine. Les dispositions n'étant accessibles au sujet qu'à travers leurs manifestations, sur lesquelles il fonde ses concepts, une analyse FA doit se prononcer sur la nature de ces manifestations.

Le défenseur d'une analyse conative sera bien sûr tenté de répondre que les désirs se manifestent principalement sous la forme de désirs occurrents. Mais quelle forme précise revêtent ces désirs? Il semble que nous fassions l'expérience de nos désirs lorsque la perspective (réelle ou imaginée) de leur satisfaction ou de leur frustration devient saillante. Sylvain se rend compte de son désir d'être en sécurité lorsqu'il prend peur au milieu des bois la nuit tombée. Jacques prend conscience de son désir de posséder une Ferrari lorsqu'il se réjouit en imaginant qu'il vient de toucher un héritage. Jeanne fait l'expérience de son désir d'être respectée lorsque, s'étant mise dans une situation honteuse, la frustration de ce désir lui apparaît inévitable. Cela suggère que l'expérience d'un désir prend la forme d'une émotion particulière suscitée par sa satisfaction ou sa frustration actuelle, probable ou possible.⁴ Peut-on, à partir de ces expériences plus particulières, déboucher sur l'analyse du concept de ce qui mérite d'être poursuivi? C'est possible, mais ce concept fin dépend des concepts de ces émotions particulières et des conditions dans lesquelles elles-mêmes sont méri-

tées : il doit être tiré par abstraction des concepts de ce qui mérite la peur, la joie ou la honte. Pour cette raison, il faut d'abord examiner s'il est possible d'ancrer une analyse FA des concepts évaluatifs épais dans les concepts d'attitudes émotionnelles spécifiques et, dans un second temps, étendre cette analyse aux concepts évaluatifs fins.

Cette stratégie ascendante convainc d'autant plus dès lors que l'on se pose la question suivante. Qu'est-ce qu'un sujet incapable de saisir que différentes choses méritent d'être poursuivies de différentes façons peut comprendre du concept générique de ce qui mérite d'être poursuivi? Considérez le cas similaire du concept « être coloré » : on ne voit pas comment un aveugle incapable de comprendre les manières spécifiques d'être coloré (rouge, bleu, etc.) pourrait le maîtriser. Ce constat semble s'imposer pour tous les concepts de propriétés déterminables. En effet, dans la mesure où un objet n'exemplifie la propriété déterminable (être coloré, être digne de poursuite) qu'en exemplifiant une propriété déterminée (être bleu, être admirable), un sujet qui possède le concept de la propriété déterminable doit au moins posséder les concepts de quelques propriétés déterminées.⁵ Cela suggère que les concepts de propriétés évaluatives déterminées – épaisses – sont premiers dans l'ordre d'explication caractéristique d'une analyse FA. C'est une raison supplémentaire de ne pas ancrer l'analyse dans les désirs.

Les considérations qui précèdent ne peuvent définitivement réfuter l'analyse FA des concepts évaluatifs fins au regard des attitudes génériques.⁶ Cependant, si les couples amour-haine et désir-aversion sont les seules attitudes génériques disponibles, nous avons des raisons d'examiner les analyses FA qui ciblent les concepts de propriétés évaluatives épaisses à l'aide d'attitudes plus spécifiques. Quelle forme prennent ces analyses? L'idée fondamentale est que les attitudes spécifiques pertinentes sont nos différentes réponses émotionnelles. Elle est séduisante dans la mesure où le lien de nombreux concepts de valeurs épaisses aux émotions est marqué au niveau syntaxique : on parle ainsi du honteux, de l'effrayant, de l'admirable et du joyeux. Il est donc tentant d'éclairer ces concepts de la façon suivante : un objet ou un événement est honteux, effrayant ou admirable s'il mérite la honte, la peur ou l'admiration. Ce constat explique pourquoi les analyses des concepts de valeurs épaisses focalisées sur les émotions appropriées sont couramment défendues (D'Arms & Jacobson, 2000; Gert, 2016; Tappolet, 2011). Puisque leur succès dépend de la possibilité de relever ce que nous avons appelé le défi psychologique, il faut nous prononcer à partir d'une perspective informée sur la nature des émotions. C'est pourquoi nous allons considérer les approches contemporaines des émotions les plus susceptibles de s'allier à une analyse FA.

2. FA ET CONTENU DES ÉMOTIONS

Pour une analyse FA des concepts évaluatifs épais au vu des émotions, relever le défi psychologique consiste à éviter deux écueils. Le premier, celui de la circularité, n'est évité que si les émotions peuvent être caractérisées indépendam-

ment des concepts de valeurs que l'on cherche à comprendre – que si l'on peut caractériser le concept d'amusement sans faire appel à celui de comique, ou celui de colère sans celui d'offense. Cette section présente des théories contemporaines des émotions qui s'allient naturellement à une analyse FA, car elles s'articulent autour de l'idée selon laquelle les émotions touchent nos valeurs épaisses. Peut-on adopter l'une de ces théories et relever le défi psychologique? Nous allons suggérer que c'est impossible si l'on situe cette relation à l'horizon du *contenu représentationnel* des émotions.

On se souvient qu'une analyse FA ne peut identifier l'approbation et la désapprobation à des jugements évaluatifs : ces jugements impliquant de déployer les concepts de valeurs pertinents, ils présupposent ce que l'on cherche à expliquer. Pour la même raison, le cognitivisme classique en philosophie des émotions – avoir peur, c'est juger que quelque chose est dangereux; s'amuser, juger que quelque chose est cocasse, etc. (Nussbaum, 2003; Solomon, 1976) n'est pas une option. La stratégie qui s'impose alors consiste à comprendre les émotions comme des états psychologiques qui ne présupposent pas le déploiement de ces concepts. On rejoint ici une préoccupation au cœur des théories perceptuelles des émotions : appliquer à ces dernières l'idée courante selon laquelle les expériences perceptives possèdent un contenu non conceptuel (par exemple Crane, 1992).

Une manière naturelle de comprendre cette référence au contenu évaluatif non conceptuel revient à affirmer qu'il caractérise la façon dont les émotions se présentent à celui qui les ressent. Voir une surface rouge, c'est être frappé visuellement par sa rougeur sans que cela requière la capacité à y penser comme rouge ou comme colorée. De même, s'amuser à une bonne blague reviendrait à être frappé émotionnellement par son caractère comique, s'énervé à une remarque à être frappé par son caractère offensant (Tappolet, 2000; Wringer, 2014). À l'instar de son ancêtre, le cognitivisme classique, la théorie perceptuelle admet que les émotions ont une intentionnalité évaluative. Mais elle possède sur lui un double avantage. D'une part, en soulignant le caractère non conceptuel du format de représentation, elle peut attribuer des émotions aux animaux et aux jeunes enfants (Roberts, 1996; Tappolet, 2000). D'autre part, elle met l'accent sur la dimension phénoménologique des émotions : ce sont des expériences perceptives de valeurs qui peuvent causer et justifier des jugements évaluatifs (Döring, 2007; Tappolet, 2000).

Pour le défenseur d'une analyse FA, adopter une théorie perceptuelle est un pas dans la bonne direction, puisque ressentir une émotion ne présuppose alors plus le déploiement des concepts évaluatifs qu'il cherche à analyser. Cependant, nous ne nous intéressons pas ici directement aux émotions, mais aux *concepts* d'émotions : le but de l'analyse FA conceptuelle est d'analyser les concepts évaluatifs décrivant les émotions. En admettant que les émotions sont des perceptions de valeurs, il faut donc se demander quels concepts d'émotions un sujet peut fonder sur elles. Si les émotions se donnent à lui comme des expériences de valeurs, la réponse est à première vue que le concept d'une émotion donnée est celui de

l'expérience perceptive d'une valeur. Prendre au sérieux le modèle de la perception, c'est admettre que les valeurs se présentent au sujet qui ressent une émotion d'une façon analogue à celle dont les propriétés perçues se présentent au sujet qui perçoit.

Il a souvent été observé que décrire les expériences perceptives de la perspective du sujet revient à décrire ce qu'elles représentent. Décrire l'expérience perceptive d'un vase bleu reviendrait à décrire les propriétés spatiales et chromatiques de ce vase du point de vue occupé par le sujet. Pour cette raison, les expériences perceptives sont couramment décrites comme transparentes (Harman, 1990; Martin, 2002). Et, si elles sont transparentes, celui qui fonde un concept sur ses expériences – qui fonde par exemple le concept d'un type d'expériences sur les expériences visuelles de bleu – ne peut les concevoir qu'en tant que propriétés qu'elles lui présentent : c'est après tout la seule chose à laquelle il ait accès. Cette conclusion est difficile à réconcilier avec une analyse FA des concepts évaluatifs. Prendre le modèle de la perception au sérieux et admettre la notion de transparence qu'il véhicule, c'est conclure que le concept d'amusement est identique au concept de perception du comique, celui de colère identique à celui de perception de l'offensant. Une analyse FA alliée à une approche perceptive des émotions n'est donc pas mieux placée pour relever le défi psychologique que le cognitivisme classique; l'analyse demeure circulaire, l'échéance ayant simplement été retardée.

Toutefois, il ne faudrait pas tirer cette conclusion trop hâtivement : le défenseur d'une théorie perceptuelle peut contourner la difficulté en insistant sur le fait que la description « perception de telle ou telle valeur » n'est pas fidèle à la façon dont les émotions se donnent à celui qui les ressent. Cette stratégie a pour but de montrer que les concepts d'émotions fondés sur les expériences émotionnelles ne sont pas les concepts de perceptions de telles ou telles valeurs. Sans prétendre à l'exhaustivité, nous allons considérer deux options contemporaines : la première fait appel à la représentation de nos préoccupations, la seconde s'inspire de James.

La première option, défendue par Robert Roberts (2003), consiste à comprendre l'idée de « perception de valeur » à titre de représentation d'une situation comme satisfaisant ou frustrant les préoccupations (*concerns*) du sujet. Un sujet en colère n'est pas frappé par l'offense d'une remarque, il comprend plutôt qu'il se trouve dans une situation qui menace son souci de préserver son honneur. Bien sûr, puisque l'on se situe dans le cadre d'une théorie perceptuelle, la notion de compréhension en jeu n'implique pas la croyance. L'idée est que le sujet comprend la situation comme menaçant sa préoccupation d'une façon analogue à celle dont il peut voir un objet comme un objet d'un certain type (voir un métal comme de l'or) : il s'agit d'une compréhension souvent automatique, immédiate et compatible avec le jugement que l'objet n'est pas de ce type. Si la perception de valeur est la perception d'un rapport entre une situation et une préoccupation, nous entrevoyons à nouveau la possibilité d'une analyse non circulaire des concepts évaluatifs décrivant les émotions.

Cette approche peut-elle relever le défi psychologique? Une chose est claire, elle évite aisément l'écueil de *l'informativité*. Les attitudes retenues sont en effet assez proches de notre saisie intuitive des valeurs pour contribuer à leur compréhension : le concept de valeur semble bel et bien lié au fait que nous nous concevons comme des créatures animées de préoccupations que le cours des choses favorise ou menace. Qu'en est-il de l'écueil de la *circularité*? Il est évité si l'approche nous convainc que les émotions sont suffisamment indépendantes des valeurs pour que les concepts d'émotions ne présupposent pas ceux des valeurs pertinentes. Si, comme elle l'affirme, les émotions se présentent au sujet comme des représentations de préoccupations satisfaites ou frustrées, peut-il développer des concepts d'émotions sans se référer à ces valeurs? La réponse doit être nuancée, car elle dépend des préoccupations en jeu. Il ne faut bien sûr pas associer la peur à la préoccupation d'*éviter le danger*, la colère à celle d'*éviter les offenses*, et ainsi de suite – faute de quoi l'analyse FA reste circulaire. Il faudrait plutôt les associer à des préoccupations comme celle pour *la sécurité* (peur) et *l'honneur* (colère).

Dans la mesure où l'analyse FA du concept de danger ou d'offense est fondée sur les concepts de peur ou de colère, l'approche est attrayante : les concepts de peur et de colère sont structurés autour de la représentation de ces préoccupations pour la sécurité et l'honneur dont on admet qu'elles sont indépendantes du danger et de l'offense.⁷ Soit. Mais remarquez que les préoccupations en jeu dans ces émotions sont évaluatives : bien que l'on puisse admettre que la sécurité et l'honneur sont des valeurs indépendantes du danger et de l'offense, elles n'en demeurent pas moins des valeurs. La conséquence immédiate est que cette approche spéculé à la baisse sur la possibilité pour une analyse FA d'éclairer le domaine évaluatif. En effet, elle revient à admettre que les sujets qui développent les concepts d'émotions pertinents possèdent certains concepts évaluatifs qui ne doivent rien à leurs concepts d'émotions. C'est la seule façon d'éviter la circularité. La promesse au cœur des analyses FA n'est alors pas tenue : nos concepts d'émotions découlant d'un ensemble indépendamment constitué de concepts évaluatifs, ils ne peuvent en constituer la source. Pour ce que nous dit l'analyse FA résultant de cette approche, les concepts évaluatifs restent aussi mystérieux qu'au départ de l'enquête. Interpréter « perception d'une valeur » comme « perception de la frustration ou de la satisfaction d'une préoccupation » ne constitue donc pas une manière satisfaisante d'élaborer une théorie perceptuelle dans le cadre d'une analyse FA, car elle ne relève pas le défi psychologique.

La seconde manière de comprendre la notion de perception de valeur est constituée par l'approche perceptuelle néo-jamésienne développée par Jesse Prinz (2004). Les émotions y sont conçues comme des ressentis corporels suscités par la présence de propriétés évaluatives. L'amusement est le ressenti de changements corporels déclenchés par des situations comiques, la colère celui de changements déclenchés par des situations offensantes, et ainsi de suite. Certains de ces liens ont une base strictement biologique, alors que d'autres se mettent en place au cours du développement et à travers un processus d'acculturation. Au

sein de cette approche, comment le lien des émotions aux valeurs est-il pris en compte? Ce lien est clairement indirect : le sujet est directement conscient de changements dans son corps et, indirectement, de leur cause, à savoir la valeur qui suscite son émotion. Du point de vue du sujet, seuls les changements corporels sont manifestes, la valeur ne figurant qu'au sein d'un jugement qui établit un lien entre les ressentis corporels et la situation évaluative qui les cause. Ce n'est que par le truchement d'un tel jugement, qui est facultatif, qu'un sujet dans une situation émotionnelle prend une perspective évaluative (Deonna & Teroni, 2012).

Cette approche peut-elle relever le défi psychologique? Puisque les expériences émotionnelles sont suscitées par les valeurs sans les présenter, elle paraît idéalement placée pour éviter l'écueil de la *circularité*. En effet, les concepts d'émotions qu'un sujet fonde sur ses émotions s'articulent autour des propriétés qu'elles lui présentent directement, à savoir des configurations de changements corporels. Le concept d'une telle configuration ne présupposant pas celui d'une valeur, il est tentant d'allier l'analyse FA à cette approche néo-jamesienne. La situation est moins claire en ce qui concerne l'écueil de l'*informativité*. Les changements corporels ne sont en effet pas liés de manière évidente à notre saisie intuitive des valeurs – peuvent-ils néanmoins constituer la source de notre compréhension des concepts évaluatifs? Il est permis d'en douter. Ressentir l'accélération de son rythme cardiaque, une montée de sueur et autres contractions musculaires sont certes des états intentionnels. Ils ne sont toutefois pas dirigés vers le porteur de la valeur épaisse en jeu, mais vers notre corps, qui n'est pas ce que nous considérons dangereux dans un épisode de peur ou offensant dans un épisode de colère. Si l'on souscrit à l'approche néo-jamesienne, les émotions ne sont pas des états psychologiques qu'un sujet conçoit comme mérités par certaines situations.

Cette conclusion s'impose-t-elle aussi rapidement? Après tout, le néo-jamésien insiste sur le fait que les valeurs, si elles ne sont pas les objets immédiats des émotions, en sont les objets médiats. Admettons.⁸ Il n'en demeure pas moins que ce qui s'impose au sujet qui ressent une émotion est une suite de changements corporels au sein de laquelle la valeur ne figure pas. Cela ne signifie pas que le sujet ne peut fonder un concept de cette valeur sur cette émotion, mais que ce concept sera celui d'une cause des changements corporels qu'il ressent. C'est là l'unique contribution possible des émotions aux concepts évaluatifs. Or, ce constat débouche immédiatement sur un dilemme. D'un côté, on peut voir dans cette conception des valeurs comme « causes inconnues des changements corporels ressentis » une élucidation des concepts de valeurs épaisses. Mais l'admettre revient à sortir du cadre de l'analyse FA pour embrasser une analyse purement dispositionnelle de ces concepts.⁹ De l'autre côté, on peut vouloir rester dans le cadre de l'analyse FA, mais on ne voit pas alors comment le concept de cause de changements corporels élucide nos concepts évaluatifs. Encore une fois, cela ne revient pas à nier que le défenseur d'une approche néo-jamesienne puisse insister sur le fait que les émotions, en tant que changements corporels ressentis, signalent la présence de valeurs. Nous cherchons plutôt à souligner que rien

dans ces signaux ne constitue une compréhension des valeurs en question : c'est l'une des raisons pour lesquelles nos concepts de valeurs épaisses ne font pas référence à des changements corporels de ce type (Deonna & Teroni, 2012).

Dans cette section, nous avons vu que les approches contemporaines selon lesquelles les émotions ont un *contenu évaluatif* ne parviennent pas à relever le défi psychologique. Partant du constat qu'une analyse FA ne peut pas s'allier à une théorie perceptuelle selon laquelle les émotions se présentent au sujet comme des perceptions de valeurs (sans quoi on échoue sur l'écueil de la circularité), nous avons passé en revue deux façons de ne pas prendre la description « perception de valeur » pour argent comptant. Nous avons conclu que l'approche de Roberts débouche sur une analyse FA non généralisable qui perd beaucoup de son attrait. Et, bien que l'approche néo-jamésienne contourne l'écueil de la circularité, elle repose sur une conception des émotions au sein de laquelle leur rôle dans la compréhension des valeurs épaisses est mystérieux. Les approches considérant le contenu évaluatif s'allient donc malaisément à une analyse FA.

3. FA ET ATTITUDES ÉMOTIONNELLES

Cette conclusion doit nous encourager à explorer une approche alternative, selon laquelle les émotions ne possèdent pas un contenu évaluatif (elles ne représentent pas les valeurs), mais sont des types spécifiques d'*attitudes* évaluatives. Après tout, concevoir des émotions telles que la honte, la peur et l'admiration comme différentes attitudes plutôt qu'une seule et même attitude – par exemple, celle de juger ou de percevoir – dirigée vers différentes valeurs semble intuitivement plausible. N'est-il pas évident que ces trois émotions sont trois attitudes distinctes que nous prenons à l'égard de quelque chose? Le contraste entre ces émotions n'est-il pas similaire à celui entre désirer, croire et conjecturer, et foncièrement distinct de celui entre croire une proposition donnée et en croire une autre? Selon nous, des réponses positives à ces deux questions s'imposent : la différence entre s'émouvoir, juger et percevoir, de même que celle entre différents types d'émotions, est avant tout une différence des attitudes prises à l'égard d'un contenu et non une différence de contenu.¹⁰

Bien qu'elle soit intuitivement plausible, une approche des liens entre émotions et valeurs à travers les attitudes plutôt que les contenus dépend naturellement d'une analyse convaincante des attitudes en jeu – analyse qui doit de surcroît, dans le contexte de notre discussion, relever le défi psychologique. Que sont ces attitudes émotionnelles? Nous pensons que l'approche néo-jamésienne abordée dans la section 2 insiste à raison sur la dimension incarnée des émotions. En revanche, elle la décrit d'une façon qui n'est pas fidèle au rôle du corps dans la conscience émotionnelle. La stratégie qui s'impose consiste à rejeter l'idée selon laquelle le ressenti corporel est une suite de sensations localisées dans diverses parties du corps. L'effet produit lorsque l'on ressent les changements corporels caractéristiques d'une émotion doit plutôt être compris de façon globale ou holistique comme l'expérience d'une prise de position, d'une posture ou d'une atti-

tude par rapport à une situation se déroulant au-delà du corps.¹¹ Cet appel à une conscience du corps plus globale et dirigée vers l'extérieur s'oppose à une conception atomiste du ressenti corporel (Claparède, 1928; Deonna & Teroni, 2012, à paraître).

Une manière séduisante de poursuivre ces réflexions consiste à recourir à un aspect des émotions souvent souligné par les approches contemporaines : les tendances à l'action (Frijda 1986, 2007; Scarantino, 2014). Parler d'expérience de tendances à agir permet en effet de développer de manière intéressante l'idée selon laquelle les émotions, en tant qu'attitudes corporelles ressenties, sont liées aux valeurs. Nous parlons ici de tendance à l'action au sens large, afin d'inclure les tendances à s'éloigner, se rapprocher ou aller à l'encontre d'un objet, à se focaliser sur lui, s'y soumettre, être attiré par lui, s'en désengager ou même suspendre toute interaction avec lui. Ainsi, lors de la peur nous ressentons notre corps se mobiliser afin de neutraliser quelque chose; lors de la colère, nous le sentons se préparer à une forme d'hostilité active; lors d'un épisode de honte, nous ressentons une promptitude à nous éloigner du regard de celui qui cause cette émotion; et, lorsque nous sommes tristes, nous sentons que notre corps est privé de la possibilité d'interagir avec l'objet dont la perte suscite cette émotion.¹²

Ces exemples qui soulignent la dimension agentielle du ressenti propre aux émotions nous permettent de comprendre leur lien aux valeurs. La peur d'un chien est une attitude évaluative, dans la mesure où elle consiste en l'expérience du corps comme préparé à limiter son impact (fuite, attaque préventive, immobilité, etc.), une attitude qu'il est correct d'adopter si, et seulement si, le chien est dangereux. La colère vis-à-vis d'une personne est une autre attitude évaluative, dans la mesure où elle est une expérience du corps comme préparé à la revanche, une attitude correcte si, et seulement si, la personne est offensante. La honte est une troisième attitude évaluative dans laquelle nous ressentons notre corps comme prêt à disparaître hors de vue de ceux qui nous dénigrent, une attitude correcte lorsque nous sommes effectivement dans une situation dégradante. Ce sont là quelques exemples de l'approche alternative du lien entre émotions et valeurs qui nous souhaitons promouvoir. Elle se démarque des conceptions selon le contenu évaluatif discutées dans la section précédente en rejetant l'idée selon laquelle les valeurs figurent dans le contenu des émotions. Dans cette mesure, elle pourrait bien s'allier à une analyse FA des concepts évaluatifs et relever le défi psychologique.

Pour rappel, relever ce défi revient à expliquer comment un sujet peut acquérir le concept d'une valeur épaisse à partir du concept d'une émotion donnée tout en évitant deux écueils : le concept d'émotion ne doit pas présupposer celui de la valeur (*circularité*), tout en demeurant le concept d'une attitude susceptible d'éclairer notre compréhension des valeurs (*informativité*). Il nous semble que la théorie attitudinale est bien placée pour les éviter. D'abord, elle a le même avantage que la théorie néo-jamésienne : tout comme le concept d'une suite de changements corporels, le concept d'une tendance ressentie à agir sur l'envi-

ronnement de telle ou telle façon ne semble pas présupposer le concept d'une quelconque valeur. Ensuite, un tel concept peut éclairer notre compréhension des valeurs : n'est-il pas plausible d'affirmer que le cœur des concepts de danger, d'offense et de l'admirable consiste en notre compréhension que certains objets ou événements méritent, respectivement, une certaine forme de neutralisation, d'hostilité et d'attention prospectrice? À notre avis, la dimension agentielle de ces descriptions capture d'une façon convaincante la spécificité des concepts évaluatifs épais, ce qui nous encourage à donner une réponse positive à cette question.

La théorie attitudinale esquissée ici semble donc en mesure de relever le défi psychologique. Au contraire des options discutées plus haut, elle gagne le droit de relever le défi normatif.

4. DU PSYCHOLOGIQUE AU NORMATIF

Le défi normatif auquel fait face une analyse FA des concepts évaluatifs provient du fait qu'elle fait appel à des attitudes *appropriées* ou *méritées*. Il y a donc deux nouveaux écueils à éviter. D'un côté, le concept d'une émotion appropriée ne doit pas faire appel au concept de la valeur pertinente – c'est la version normative de l'écueil de la circularité. De l'autre, il ne faut pas que la notion de l'approprié en jeu soit totalement déconnectée de la nature des émotions, sans quoi on ne comprendrait plus la raison d'un détour par celles-ci – c'est la version normative de l'écueil de l'informativité.

Considérons d'abord la question de la *circularité*, qui concerne toute conception des émotions que l'on peut allier à une analyse FA. Le concept de mérite pour une émotion donnée présuppose-t-il le concept de la valeur pertinente? Peut-on comprendre ce qu'est un épisode de colère mérité sans comprendre ce qu'est une offense? Nous sommes d'avis que oui. La clé consiste à faire appel, non pas directement à la propriété évaluative, mais plutôt aux propriétés naturelles qui, dans des circonstances données, constituent cette propriété évaluative. Prenons un cas simple : un sujet prend peur à la vue d'un chien s'approchant de lui. Sa peur sera appropriée s'il est conscient de propriétés (grognements, dents acérées, poils hérissés, queue dressée) qui, dans les circonstances, constituent un danger, et prend peur parce qu'il est conscient de ces propriétés (Deonna & Teroni, 2012; Teroni 2013). L'idée consiste à considérer ces cas particuliers comme premiers dans la genèse du concept d'émotion appropriée : ce concept est, dans le cas qui nous occupe, celui d'une constellation de propriétés naturelles données, dans un autre cas de peur, celui d'une constellation différente de telles propriétés, et ainsi de suite. Ces différents concepts, s'ils peuvent converger vers celui de danger, ne présupposent pas son déploiement. Cette conception prometteuse du caractère mérité des émotions est susceptible d'être recrutée par les partisans d'une analyse FA.¹³

On peut objecter à cette conception des émotions méritées à titre de propriétés naturelles le fait suivant : comprendre ce qu'est une émotion méritée présup-

pose que l'on saisisse que les propriétés naturelles en question sont une raison de s'émouvoir parce qu'elles constituent une valeur. Par exemple, un sujet ne conçoit sa peur comme méritée que s'il saisit le comportement d'un chien sous l'aspect du danger. En un mot, le concept d'une émotion méritée serait le concept d'une émotion suscitée par des propriétés naturelles *appréhendées sous un jour évaluatif*. Cette objection doit-elle nous faire renoncer à la stratégie esquissée ci-dessus? Dans la mesure où elle repose sur des affirmations épistémologiques très générales, nous devons nous contenter ici d'un début de réponse. Remarquez que l'objection applique au cas qui nous intéresse une intuition défendue par certaines formes d'internalisme : selon cette intuition, les considérations qui justifient les états psychologiques d'un sujet doivent être conçues par ce sujet comme des raisons pour ces états.¹⁴ La conception des émotions méritée que nous avançons est en porte-à-faux avec cette intuition, mais cela constitue peut-être un atout. L'intuition internaliste est non seulement battue en brèche par de nombreuses approches (Alston, 1989; Bergmann, 2006), elle est aussi incompatible avec une analyse FA des concepts évaluatifs. En effet, souscrire à cette intuition revient à dire que les concepts d'émotions méritées sont ancrés dans les concepts évaluatifs pertinents et non, à l'instar d'une analyse FA, que les concepts évaluatifs sont ancrés dans les concepts d'émotions méritées. Éviter l'écueil de la circularité normative demande donc que l'on explique les concepts d'émotions méritées par les concepts des propriétés naturelles pertinentes.

Qu'en est-il de l'écueil de l'*informativité* dans sa version normative? Il est évité si l'on montre que notre compréhension du type de mérite que nous appliquons aux émotions dépend de leur nature, sans quoi il n'y aurait plus de raison de faire appel aux émotions pour comprendre cette notion au cœur de l'analyse FA. Considérez l'analogie suivante. On distingue parfois les règles d'étiquette des règles morales. La règle d'étiquette « les fourchettes doivent être placées à gauche des assiettes » fixe ce qui est approprié lorsque l'on dresse une table, mais rien dans la nature des fourchettes et des assiettes ne contraint cette pratique. Notre compréhension des couverts n'exerçant aucune contrainte sur cette règle, celui qui cherche à comprendre la nature de l'approprié en matière d'étiquette serait mal avisé de les étudier. En revanche, celui qui cherche à comprendre la règle morale « il ne faut pas battre autrui sans raison » serait bien avisé d'étudier les êtres humains, car c'est dans leur capacité à souffrir que cette règle trouve son fondement – c'est pourquoi elle ne s'applique pas aux minéraux ou aux robots.

Les raisons de s'émouvoir dont nous avons parlé sont-elles liées aux émotions d'une manière assimilable aux règles d'étiquette – règles qui s'appliquent à des entités indépendamment de leur nature, ou aux règles morales, qui s'y appliquent en vertu de leur nature? L'analyse FA doit évidemment opter pour la seconde branche de cette alternative : elle cherche à élucider les concepts évaluatifs par les concepts d'émotions méritées et présuppose que la notion de mérite est étroitement liée à notre compréhension des émotions. Il convient donc de s'interroger sur la notion de mérite en jeu et sur ses liens avec les émotions. Que veut-on dire lorsque l'on affirme que certaines situations méritent de la peur, de la colère

ou de l'admiration? Voici une réponse attrayante : on veut dire que ces situations méritent certaines formes d'évitement, de rétribution ou d'attention soutenue. C'est de cette façon que s'éclairent nos concepts évaluatifs épais. Or, on notera avec satisfaction que cette réponse fait référence aux types d'interactions privilégiés par la théorie attitudinale discutée dans la section 3 : selon cette théorie, il s'agit là des interactions auxquelles les émotions nous préparent et dont nous faisons l'expérience lorsque nous les ressentons. Le concept de mérite en jeu pour un certain type d'émotion est donc bel et bien informé par la nature de cette émotion.

Est-ce suffisant pour éviter le second écueil? Peut-être pas. Après tout, l'analyse FA promet d'élucider les concepts évaluatifs au titre de concepts d'émotions méritées et la nécessité d'un détour par les émotions n'a pas été clairement établie. Plus précisément, notre compréhension des considérations que nous employons pour évaluer si une émotion est méritée dépend-elle vraiment des réponses émotionnelles? Le cas de la peur pourrait motiver une position sceptique à ce propos : une raison d'avoir peur est constituée par ce qui menace notre intégrité physique, concept qui ne dépend pas clairement de celui de peur. Si ce cas est représentatif, le détour par les émotions est entièrement superflu et l'analyse FA ne saurait tenir sa promesse. Cependant, le concept de danger n'est pas très représentatif. Considérez les raisons de s'énervé, de s'attrister ou de s'amuser. Peut-on comprendre ces concepts sans les ancrer dans les émotions en question? Selon le défenseur de l'analyse FA, ce ne l'est pas : notre compréhension de ces raisons est en réalité fondée sur les réponses émotionnelles suscitées par des situations paradigmatiques (de Sousa, 1987; d'Arms & Jacobson, 2005). L'idée est que nous possédons une capacité plus ou moins innée à répondre par des émotions données à certaines situations et que notre concept de raisons pour ces émotions se développe à partir de ces situations. Pensez à la genèse du sens de l'offense, partant de réactions à des situations paradigmatiques qui constituent autant d'empiètements injustifiés sur les fins poursuivies par l'agent.e (bousculades, invasion de l'espace privé, etc.), embrassant ensuite des situations plus complexes (devoir longuement patienter lors d'un rendez-vous) pour s'étendre enfin à des événements distants (scandales politiques à l'étranger). Or, si notre sensibilité affective est susceptible de se développer de différentes façons, sa plasticité est limitée : c'est uniquement lorsqu'un sujet perçoit une similitude entre une situation nouvelle et ces situations paradigmatiques qu'il étend sa capacité à ressentir une émotion donnée. Dans notre exemple, le développement du sens de l'offense s'explique par la saisie de similitudes entre les bousculades, les attentes et les scandales politiques, qui sont autant d'atteintes injustifiées aux fins des agents concernés. Si ces considérations sur la colère sont généralisables, cela suggère que les concepts de raisons pour les différentes émotions dépendent essentiellement de l'ancrage de ces émotions dans certaines situations paradigmatiques et de la plasticité limitée de notre sensibilité. C'est pourquoi notre compréhension de ces raisons est essentiellement liée à ces émotions.

Enfin, on pourrait concéder qu'une analyse FA alliée à la théorie attitudinale des émotions est capable d'éviter ces deux écueils normatifs, mais au prix d'un amal-

game entre différents types de normativité.¹⁵ Nous avons souligné que la saisie de l'offense est essentiellement la saisie de quelque chose qui mérite un certain type d'hostilité. S'agit-il là d'un mérite proprement épistémique, ou de quelque chose de nature plutôt prudentielle ou morale (Smith, 2014)? Une offense mérite-t-elle au sens épistémique de l'hostilité, plutôt que de « tendre l'autre joue »?

Ces questions difficiles ne peuvent recevoir ici les réponses détaillées qu'elles méritent. Nous pouvons cependant esquisser quelques pistes de réflexion. La première remarque qui s'impose est que, pour de nombreuses émotions, la démarcation entre ce qui est épistémique, prudentiel et moral n'est pas aussi tranchée que l'on pourrait le penser. La peur et la tristesse sont ainsi correctes s'il y a danger et perte, deux valeurs prudentielles. Et ce sont les fautes morales qui méritent le remords. Il ne faut donc pas perdre de vue que la notion de mérite, aussi épistémique soit-elle, relie souvent une attitude psychologique à une valeur prudentielle ou morale. Cette conclusion implique-t-elle que différents types de raisons ne peuvent pas être distingués? Certainement pas : elle implique seulement que les raisons épistémiques pour les émotions coïncident parfois avec des raisons prudentielles ou morales. Ainsi, les raisons épistémiques de se mettre en colère trouvent leur source dans une offense. L'offense étant elle-même une valeur prudentielle (ou peut-être morale), ces raisons épistémiques sont autant de raisons de réagir à la présence d'une telle valeur. Pour autant, ces différents types de raisons ne coïncident pas parfaitement. Par exemple, les raisons épistémiques de ressentir de la colère sont des raisons prudentielles *pro tanto* : toutes choses considérées, l'offense particulière et les circonstances qui l'entourent préconisent parfois de « tendre l'autre joue ». En outre, certaines raisons prudentielles de s'énerver n'ont rien à voir avec l'offense – la perception de la colère du sujet par autrui peut ainsi lui éviter d'être maltraité dans le futur. Il n'est donc pas besoin de nier qu'il existe différents types de raisons pour les émotions.

Nous pensons avoir montré que la crainte selon laquelle l'approche développée ici amalgame différents types de raisons est sans fondement. Elle demeurera cependant chez ceux qui voient d'un mauvais œil le fait d'allier l'analyse FA à une conception des émotions selon laquelle elles sont orientées vers l'action. La discorde proviendrait alors d'un présupposé selon lequel la notion de mérite épistémique n'a de sens que pour des attitudes qui, à l'instar de la perception, présentent la réalité d'une façon neutre ou non engagée. Nous espérons que la discussion des théories se focalisant sur le contenu représentationnel des émotions aura contribué à battre ce présupposé en brèche. En effet, une façon de résumer la leçon que l'on peut tirer de notre discussion est que, si le modèle de la représentation convient à certains domaines, comme celui des couleurs, ce sont précisément ceux dans lesquels la notion de mérite ne s'applique pas.¹⁶ Le fait que le concept de bleu n'est pas un concept évaluatif est reflété dans le fait qu'il ne s'agit pas du concept de quelque chose qui mérite une réponse affective. La spécificité des concepts évaluatifs réside dans leur référence à ce qui mérite une certaine forme d'engagement.

NOTES

- ¹ Franz Brentano (1889/1969) est souvent considéré comme le père de ces approches et Ewing (1947, 1959) comme l'un de ses avocats les plus notoires. McDowell (1985) et Scanlon (1998) en sont des défenseurs récents et influents. Une histoire détaillée et une discussion des déclinaisons possibles de l'analyse FA se trouvent dans Rabinowicz et Rønnow-Rasmussen (2004). Pour un examen critique approfondi, voir Louise (2009) et Gert (2015).
- ² Certaines approches des rapports entre (concepts de) valeurs et (concepts d') émotions ne considèrent pas la circularité comme un problème, car elles reposent sur une forme de holisme (Helm, 2001) ou sur l'idée que le but d'une analyse de type FA consiste à éclairer nos concepts évaluatifs plutôt qu'à en offrir une analyse au sens strict (Wiggins, 1987). Rien dans la discussion qui suit ne va à l'encontre de ces approches. Nous estimons toutefois qu'il est intéressant en soi d'examiner si une analyse non circulaire est défendable. Si aucune ne l'est, cela constituerait en effet une motivation supplémentaire de développer une approche à circularité vertueuse.
- ³ La distinction entre valeurs fines et épaisses est introduite dans la littérature contemporaine par Williams (1985). Pour une discussion récente, voir Väyrynen (2013).
- ⁴ C'est ce qui explique la popularité des conceptions des émotions selon lesquelles elles représentent la satisfaction ou la frustration de nos désirs. Voir par exemple Wollheim (1999) et Schroeder (2004).
- ⁵ La distinction entre propriétés déterminables et déterminées est présentée de façon systématique par Johnson (1921). Pour une discussion récente, voir Wilson (2017).
- ⁶ On pourrait en effet chercher à montrer deux choses. D'une part, que les attitudes génériques peuvent se manifester en tant que telles indépendamment d'attitudes spécifiques. D'autre part, que les concepts de valeurs fines ne sont pas dans une relation déterminable-déterminé avec les concepts de valeurs épaisses et donc qu'une caractérisation des valeurs fines informative et indépendante des valeurs épaisses est disponible au sujet.
- ⁷ On peut étayer cette indépendance en observant que les préoccupations en question sont suffisamment génériques pour ne pas correspondre à des valeurs uniques. Ainsi, l'honneur est en jeu dans des situations offensantes (colère) tout autant que dégradantes (honte).
- ⁸ Cette concession implique que l'on accepte le cadre théorique de la sémantique informationnelle (Dretske et Millikan, 1984) dans lequel s'inscrit l'approche en question.
- ⁹ C'est d'ailleurs l'approche défendue par Prinz (2007) lui-même.
- ¹⁰ Nous donnons des raisons d'adopter une telle théorie des émotions dans la perspective des attitudes dans Deonna & Teroni (2012, 2015 et 2016).
- ¹¹ Observez que les émotions n'ont pas pour objet le corps ressenti, mais qu'il s'agit selon l'approche en question de ressentis du corps comme engagé ou mobilisé d'une certaine façon vis-à-vis de l'environnement. Le « de » dont il est question lorsque l'on parle de « ressenti *de* son corps comme engagé » spécifie le type de ressenti et n'implique pas que le corps soit l'objet intentionnel de l'attitude. Sur la distinction entre le « de » de la spécification et celui de l'intentionnalité, voir Massin (2013), Deonna & Teroni (à paraître) et Teroni (2017).
- ¹² Un lectorat bien disposé remarquera dans ces descriptions une extension potentiellement attrayante de la notion d'attitude corporelle ressentie à des activités avant tout liées à l'attention.
- ¹³ On retrouve cette stratégie faisant appel aux propriétés naturelles dans les approches contemporaines des raisons développées indépendamment par Dancy (2004), Scanlon (1998) et, dans le domaine affectif, par Goldie (2004). Voir aussi Setiya (2012).
- ¹⁴ Nous développons ici l'intuition internaliste sous l'angle des jugements évaluatifs. Une autre méthode consiste à faire appel à des sources d'information évaluative de plus bas niveau, comme des expériences de valeurs. Nous critiquons ailleurs cette méthode (Deonna & Teroni, 2012), dont on peut par ailleurs douter qu'elle réponde à l'intuition considérée ici : faire l'expérience d'une valeur et y répondre par une émotion, ce n'est pas encore considérer la valeur comme une raison de ressentir cette émotion.

¹⁵La discussion qui suit est liée au problème du « mauvais type de raisons » (Rabinowicz & Rønnow-Rasmussen, 2004), aussi connu comme le « problème de l'amalgame » (D'Arms & Jacobson, 2000). Ce problème consiste à souligner qu'il existe de nombreuses raisons de natures différentes pour ressentir des émotions – il incombe donc au tenant de l'analyse FA d'expliquer comment il sélectionne parmi ces raisons celles qui sont pertinentes pour l'analyse des valeurs.

¹⁶Cette distinction entre concepts de couleurs et concepts évaluatifs a souvent été soulignée par les tenants d'une analyse FA, comme McDowell (1985).

BIBLIOGRAPHIE

- Alston, W., *Epistemic Justification*, Cornell, Cornell University Press, 1989.
- Bergmann, M., *Justification Without Awareness. A Defense of Epistemic Externalism*, New York, Oxford University Press, 2006.
- Brentano, F., *The Origin of our Knowledge of Right and Wrong*, London, Routledge & Kegan Paul, 1969 [1889].
- Broad, C.D., « Emotion and Sentiment », in D. Cheeney (dir.), *Broad's Critical Essays in Moral Philosophy*, New York, George Allen & Unwin, 1954, p. 283-301.
- Crane, T., « The Nonconceptual Content of Experience », in T. Crane (dir.), *The Contents of Experience : Essays on Perception*, Cambridge, Cambridge University Press, 1992, p. 136-157.
- Dancy, J., *Ethics Without Principles*, New York, Oxford University Press, 2004.
- Deonna, J. et F. Teroni, « Taking Affective Explanations to Heart », *Social Science Information*, vol. 48, no. 3, 2009, p. 359-377.
- , *The Emotions : A Philosophical Introduction*, New York, Routledge, 2012.
- , « Emotions as Attitudes », *Dialectica*, vol. 69, no. 3, 2015, p. 293-311.
- , « Getting Bodily Feelings into Emotional Experience in the Right Way », *Emotion Review*, vol. 9, no. 1, 2016, p. 55-63.
- , « Emotional Experience », in U. Kriegel (dir.), *The Oxford Handbook of Philosophy of Consciousness*, New York, Oxford University Press (à paraître).
- Döring, S., « Affective Perception and Rational Motivatio », *Dialectica*, vol. 61, no. 3, 2007, p. 363-394.
- D'Arms, J. et D. Jacobson, « The Moralistic Fallacy : On the "Appropriateness" of Emotions », *Philosophy and Phenomenological Research*, vol. 61, no. 1, 2000, p. 65-90.
- , « Anthropocentric Constraints on Human Value », *Oxford Studies in Metaethics*, vol. 1, 2005, p. 99-126.
- De Sousa, R., *The Rationality of Emotions*, Cambridge, Mass, MIT Press, 1987.
- Dokic, J. et S. Lemaire, « Are Emotions Evaluative Modes? », *Canadian Journal of Philosophy*, vol. 43, no. 2, 2015, p. 227-247.
- Dretske, F., *Naturalizing the Mind*, Cambridge, Mass., The MIT Press, 1995.
- Ewing, A. C., *The Definition of Good*, London, Routledge & Kegan Paul, 1948.
- Frijda, N., *The Emotions*, Cambridge, Cambridge University Press, 1986.
- , *The Laws of Emotion*, Mahwah, New Jersey, Lawrence Erlbaum, 2007.

Gert, J., « A Fitting End to the Wrong Kind of Reasons Problem », *Ethics*, vol. 126, no. 4, 2016, p. 1015-1042.

Goldie, P., « Emotion, Feeling and Knowledge of the World », in R. Solomon (dir.), *Thinking about Feeling*, New York, Oxford University Press, 2004b, p. 91-106.

Harman, G., « The Intrinsic Quality of Experience », in J. E. Tomberlin (dir.), *Philosophical Perspectives 4 : Action Theory and the Philosophy of Mind*, Atascadero, CA., Ridgeview, 1990, p. 31-52.

Helm, B., *Emotional Reason : Deliberation, Motivation, and the Nature of Value*, Cambridge, Cambridge University Press., 2001.

Johnson, W., *Logic (Part 1)*, Cambridge, Cambridge University Press, 1921.

Louise, J., « Correct Responses and the Priority of the Normative », *Ethical Theory and Moral Practice*, vol. 12, no. 4, 2009, p. 345-364.

McDowell, J., « Values and Secondary Qualities », in J. McDowell, *Mind, Value, and Reality*, Cambridge, Mass., Harvard University Press, 1985.

Martin, M., « The Transparency of Experience », *Mind and Language*, vol. 4, no. 4, 2002, p. 376-425.

Massin, O., « The Intentionality of Pleasures and Other Feelings. A Brentanian Approach », in D. Fisette et G. Fréchette (dir.), *Themes from Brentano*, Amsterdam, Rodopi, 2013, p. 307-338.

Millikan, R., *Language, Thought, and Other Biological Categories : New Foundations for Realism*, Cambridge, Mass., MIT Press, 1987.

Naar, H., « A Dispositional Theory of Love », *Pacific Philosophical Quarterly*, vol 94, no. 3, 2013, p. 342-357.

Nussbaum, M., *Upheavals of Thought : The Intelligence of Emotions*, Cambridge, Cambridge University Press, 2003.

Oddie, G., *Value, Reality and Desire*, New York, Oxford University Press, 2005.

Prinz, J., *Gut Reactions : A Perceptual Theory of Emotions*, New York, Oxford University Press, 2004.

Rabinowicz, W. et Rønnow-Rasmussen, T., « The Strike of the Demon : On Fitting Pro-Attitudes and Value », *Ethics*, no. 114, p. 391-423.

Roberts, R., « Propositions and Animal Emotion », *Philosophy*, vol. 71, no. 275, 1996, p. 147-156.

—, *Emotions: An Essay in Aid of Moral Psychology*, Cambridge, Cambridge University Press, 2003.

Scanlon, T. M., *What We Owe to Each Other*, Cambridge, Mass., Harvard University Press, 1998.

Scarantino, A., « The Motivational Theory of Emotion », in D. Jacobson & J. D'Arms (dir.), *Moral Psychology and Human Agency*, New York, Cambridge University Press, 2014b, p. 156-185.

Schroeder, T., *Three Faces of Desire*, New York, Oxford University Press, 2004.

Setiya, K., *Knowing Right from Wrong*, New York, Oxford University Press, 2012.

Solomon, R., *The Passions : Emotions and the Meaning of Life*, Indianapolis, Hackett, 1993.

Tappolet, C., *Émotions et valeurs*, Paris, Presses Universitaires de France, 2000.

—, « Values and Emotions : Neo-Sentimentalism's Prospects », in C. Bagnoli (dir.), *Morality and the Emotions*, New York, Oxford University Press, 2011, p. 117-134.

Teroni, F., « Emotions et connaissance », in J.-M. Chevalier et B. Gaultier (dir.), *Connaître : Questions d'épistémologie contemporaine*, Paris, Ithaque, 2014, p. 9-34.

—, « In Pursuit of Emotional Modes : The Philosophy of Emotion After James », in A. Cohen et B. Stern (dir.), *Thinking About the Emotions : A Philosophical History*, Oxford, Oxford University Press, 2017, p. 291-313.

Väyrynen, P., *The Lewd, the Rude and the Nasty*, New York, Oxford University Press, 2013.

Wiggins, D., « A Sensible Subjectivism? », in *Needs, Values, Truth : Essays in the Philosophy of Value*, Oxford, Blackwell, 1987.

Williams, B., *Ethics and the Limits of Philosophy*, Cambridge, MA, Harvard University Press, 1985.

Wilson, J., « Determinables and Determinates », in E. Zalta (dir.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), URL = <<https://plato.stanford.edu/archives/spr2017/entries/determinate-determinables/>>.

Wollheim, R., *On the Emotions*, Yale University Press, 1999.

Wringe, B., « The Contents of Perception and the Contents of Emotion », *Noûs*, vol. 48, no. 1, 2014, p. 275-297.

LE RÉALISME MORAL

PETER RAILTON
UNIVERSITY OF MICHIGAN

TRADUIT PAR DENIS COURVILLE
CANDIDAT AU DOCTORAT EN PHILOSOPHIE, UNIVERSITÉ DU QUÉBEC À MONTRÉAL

Il existe une tendance, parmi les philosophes contemporains, à considérer le scepticisme moral comme une position irrésistible, même chez ceux pour qui le scepticisme envers les sciences empiriques ne représente absolument pas une position convaincante. Pour diverses raisons, parmi lesquelles on trouve une méfiance compréhensible envers l'absolutisme moral, on considère comme un signe de bon sens d'expliquer [*explain away*] toute apparence d'objectivité dans le domaine du discours moral. Il est devenu si commun de traiter la moralité comme subjective ou conventionnelle dans le cadre d'une culture intellectuelle séculière que la plupart d'entre nous éprouve de la difficulté à s'imaginer ce à quoi pourrait ressembler l'existence de faits auxquels les jugements moraux sont redevables.

Loin d'être découragés, certains philosophes ont tenté de rendre compte de l'objectivité de la moralité en faisant valoir que la raison ou la science sert de fondement à l'éthique. Toutefois, l'histoire des tentatives de ce type n'inspire guère la confiance. Bien que le rationalisme en éthique ait retenu certains adhérents bien après que les autres rationalismes aient été abandonnés, les courants philosophiques puissants qui ont érodé l'idée d'une raison qui pourrait être, à elle seule, à même de nous offrir une position en fonction de laquelle des conclusions substantielles pourraient être dérivées ne manifestent aucun signe de fléchissement. Qui plus est, l'éthique naturaliste n'est pas parvenue jusqu'à présent à formuler une synthèse plausible des champs empirique et normatif : plus elle s'adonne à des analyses descriptives au sujet de l'origine des normes, moins elle retient sa force morale que l'on reconnaît; plus elle tente d'offrir une base reconnaissable à la critique ou à la reconstruction morale, moins elle retient un lien solide avec les théories psychologiques ou sociales descriptives.¹

Dans ce qui suit, je présenterai de manière programmatique une forme de naturalisme éthique qui, bien que considérablement redevable aux idées des théoriciens antérieurs, cherche à établir une relation plus satisfaisante entre les champs normatif et empirique. Cette relation, je crois, ne peut être démontrée par aucune preuve. Mon objectif ne consiste pas à réfuter le scepticisme moral, tout comme l'objectif des naturalistes épistémiques contemporains ne consiste pas à réfuter le scepticisme cartésien. Les prétentions du naturaliste, dans un cas comme dans l'autre, sont plus modestes. Il tente d'abord d'offrir une analyse de l'épistémologie ou de l'éthique permettant de saisir comment les fonctions évaluatives centrales dans ces domaines pourraient être accomplies dans le cadre des théories empiriques actuelles (ou prospectives). Il tente ensuite de montrer comment les explications non naturalistes traditionnelles s'appuient sur des suppositions qui sont, d'une certaine manière, incohérentes, ou qui s'accordent mal avec l'état actuel des sciences. Troisièmement, il lance un certain défi au sceptique, en montrant notamment comment une explication sceptique de nos pratiques épistémiques et morales peut être aussi plausible, utile ou intéressante que l'explication proposée par le naturalisme, et comment une reconstruction de telles pratiques – si le sceptique devait tenter d'en offrir une, comme il lui arrive souvent de le faire – peut parvenir à conserver la place et la fonction typiques de ces pratiques dans l'ordre des questions humaines. Je me concentrerai principalement sur la première de ces trois aspirations.

Une chose doit d'emblée être affirmée. Certains peuvent être interpellés ou repoussés par le réalisme moral sous prétexte qu'il s'agit là d'une conception de l'éthique qui est le plus à même d'exprimer le sérieux d'une haute considération pour la moralité. Or la moralité peut être prise au sérieux, même de manière excessive, sans qu'elle implique le réalisme moral. En effet, une objection possible pouvant être adressée au type de réalisme moral que je défends ici maintient que cette position peut ne pas considérer la moralité avec suffisamment de sérieux.

I. LES ESPÈCES DE RÉALISME MORAL

Diverses positions ont affirmé être – ou sont accusées d'être – réalistes en matière de moralité, de sorte qu'une caractérisation initiale de la position que je défends ici est requise avant que l'on puisse poursuivre davantage l'analyse. Les affirmations – et les accusations – de réalisme moral se fondent généralement sur certaines ou l'ensemble des questions suivantes. Grosso modo : 1) Le cognitivisme – Est-ce que tous les jugements moraux sont capables de vérité et de fausseté? 2) Les théories de la vérité – Si les jugements moraux ont des valeurs de vérité, en quel sens est-ce le cas? 3) L'objectivité – De quelles façons, s'il en est, l'existence de propriétés morales dépend-elle des états d'esprit actuels ou possibles d'êtres intelligents? 4) Le réductionnisme – Les propriétés morales sont-elles réductibles à – ou surviennent-elles en un sens plus faible – des propriétés non morales? 5) Le naturalisme – Les propriétés morales sont-elles des propriétés naturelles? 6) L'empirisme – En venons-nous à une connaissance des faits moraux de la même façon dont nous en venons à connaître les faits des sciences empiriques, ou sont-ils révélés par la raison ou par un certain mode d'appréhension spécifique? 7) La bivalence – est-ce que le principe du tiers exclu s'applique aux jugements moraux? 8) La déterminabilité – Quelle que soit la procédure dont nous disposons pour évaluer les jugements moraux, quelle proportion de la moralité peut vraisemblablement être déterminable? 9) La catégoricité – Est-ce que tous les agents rationnels ont nécessairement des raisons pour obéir à des impératifs moraux? 10) L'universalité – Les impératifs moraux s'appliquent-ils à tous les agents rationnels, même – s'il en existe – à celles et ceux qui ne disposent d'aucune raison pour s'y conformer? 11) L'évaluation des modalités existantes – Est-ce que les croyances morales actuelles sont approximativement vraies, ou est-ce que les intuitions morales qui prévalent constituent, en un tout autre sens, des données privilégiées? 12) Le relativisme – Est-ce que la vérité ou la justification des jugements moraux dépend directement des normes ou des pratiques individuellement – ou socialement – adoptées? 13) Le pluralisme – existe-t-il une seule forme de la vie bonne ou un seul code moral juste (*right moral code*), ou est-ce que des formes de vie ou des codes moraux différents pourraient être appropriés dans différentes circonstances?

Voici alors les coordonnées approximatives de ma propre position au sein de cet espace conceptuel à dimensions multiples. Je veux défendre une forme de réalisme qui fait valoir que des jugements moraux peuvent être des porteurs de valeurs de vérité [*can bear truth values*] en un sens fondamentalement non épis-

témique; que les propriétés morales sont objectives bien que relationnelles; que les propriétés morales surviennent sur les propriétés naturelles et peuvent se réduire à celles-ci; que l'analyse morale relève de la recherche empirique; que l'on ne peut pas savoir à priori si la bivalence s'applique aux jugements moraux ou la façon dont de tels jugements peuvent être évalués de façon déterminante; qu'il y a des raisons de penser que nous connaissons une part importante de la moralité, mais aussi qu'il y a des raisons de croire que les conceptions morales actuelles sont fausses à certains égards et pourraient même être fausses de façon très générale; qu'un agent rationnel pourrait n'avoir aucune raison d'obéir à des impératifs moraux, bien que ceux-ci puissent néanmoins s'appliquer à elle ou à lui; et que, bien que l'évaluation morale puisse reposer sur des critères parfaitement généraux, et en vertu de la nature même de ces critères, aucun type de vie n'est susceptible de convenir à tous les individus, de même qu'aucun ensemble de normes ne peut convenir à toutes les sociétés et à toutes les époques. La position ainsi décrite pourrait bien se nommer le « réalisme moral complètement délirant » [*stark, raving moral realism*], mais pour des raisons de syntaxe, je la nommerai fadement « réalisme moral ». L'utilisation de ces expressions n'est pas exclusive à la position décrite ici. D'autres positions, qui occupent des coordonnées plus ou moins distinctes, pourraient revendiquer l'une de ces deux appellations.

II. LA DISTINCTION FAIT-VALEUR

Toute tentative visant à justifier une conception naturaliste du réalisme moral fonce tête baissée vers la distinction fait-valeur. Les diverses formulations de cette distinction, tout comme les arguments justifiant sa raison d'être, ont déjà été présentés dans le passé, mais j'insisterai, en l'occurrence, sur quelques questions concernant le statut épistémique et ontologique des jugements de valeur par opposition aux jugements de fait.

Il semble que l'argument le plus souvent évoqué en faveur de la distinction fait-valeur est d'ordre épistémique : on affirme que les conflits sur des questions de valeur peuvent persister même lorsque tous les moyens d'adjudication rationnels et scientifiques ont été déployés, de sorte que les jugements de valeur ne peuvent être cognitifs au même titre que le sont les jugements factuels ou logiques. Cette thèse est, en partie, défendue en invoquant le caractère instrumental (hypothétique) de la raison, lequel empêcherait la raison de dicter les valeurs ultimes. En principe, cet argument affirme que deux individus qui diffèrent sur des valeurs ultimes pourraient, sans faire preuve de lacunes rationnelles, tenir ferme à leurs valeurs opposées, peu importe le nombre d'arguments ou de preuves mis de l'avant. À l'instar d'Ayer, « nous estimons que la discussion est possible dans les questions de morale, seulement si un système de valeurs est présupposé. »²

On pourrait tenter de contrer cette conclusion en remettant en cause la conception instrumentale de la rationalité. Mais malgré tous ses défauts et tout ce qu'elle se doit de développer, la conception instrumentale de la rationalité semble, à mon avis, la notion la plus claire dont on dispose concernant ce que

signifient pour un agent d'avoir des raisons d'agir. De plus, elle rend compte d'une caractéristique normative centrale du fait de donner des raisons [*reason-giving*], étant donné que nous pouvons aisément reconnaître la force impérieuse de l'idée qui s'impose à l'agent, selon laquelle un acte quelconque pourrait favoriser la réalisation de ses fins. Il serait difficile de faire sens de quelqu'un qui affirmerait sincèrement à la fois avoir certaines fins et, pourtant, qu'elles ne constituent pas pour lui des raisons *prima facie* suffisantes d'agir. (Bien sûr, il pourrait également croire qu'il a d'autres raisons suffisantes, voire compensatoires).

Pourtant, cette version de l'argument épistémique en faveur de la distinction fait-valeur est aux prises avec des difficultés même si l'on concède la conception instrumentale de la rationalité. Du point de vue de la raison instrumentale, la formation de croyances n'est qu'une activité parmi d'autres : dans la mesure où nous avons des raisons de nous y consacrer ou de l'entreprendre d'une manière plutôt qu'une autre, il s'agit au fond de questions concernant leur contribution à nos fins.³ Ce qui, pour un individu, est rationnel de croire sur la base d'une quelconque expérience variera non seulement en fonction de ses autres croyances, mais aussi en fonction de ce qu'il désire.⁴ Il s'ensuit qu'aucune argumentation ou expérience, à elle seule, ne peut contraindre quelqu'un sous peine d'irrationalité d'accepter même les énoncés factuels des sciences empiriques. Le débat entamé depuis déjà très longtemps sur la logique inductive illustre bien qu'un choix rationnel parmi plusieurs hypothèses divergentes requiert des critères plus riches et plus controversés en ce qui concerne le choix d'une théorie que ceux pouvant être seulement dégagés à partir de la raison instrumentale. Malheureusement pour l'opposition qu'Ayer désirait établir, nous considérons que l'argumentation sur des questions scientifiques n'est possible que si un système de valeurs est présupposé.

Hume a toutefois trouvé un moyen d'établir la distinction entre faits et valeurs sans recourir à l'idée suivant laquelle l'induction – ou même la déduction – pourrait exiger de la part d'un agent rationnel l'adoption de certaines croyances plutôt que d'autres lorsque cela entre en conflit avec ses fins contingentes.⁵ Car Hume met de l'avant la thèse selon laquelle la moralité est pratique, à savoir que si les faits moraux existent, ils offrent nécessairement pour tout être rationnel une raison (quoique peut-être non pas une raison décisive) d'agir moralement, quoi qu'il en soit des désirs particuliers propres à chacun. Si cette thèse est retenue à titre de prémisse, la conception instrumentale de la rationalité peut être probante après tout, puisqu'elle exclut la possibilité qu'il y ait des raisons catégoriques de ce type. En revanche, Hume ne présuppose pas que cela soit constitutif de la logique ou de la science que les faits mis en évidence par celles-ci s'imposent en tant que force catégorique à tout agent rationnel, de sorte que l'existence de faits logiques ou scientifiques, à la différence de faits moraux, est compatible avec le caractère instrumental de la raison.

Or cette façon de formuler la distinction fait-valeur n'est pas aussi convaincante que la thèse suivant laquelle la moralité est essentiellement pratique au sens où

l'entend Hume.⁶ Hume a sûrement raison lorsqu'il affirme qu'il y a une relation intrinsèque, sans aucun doute complexe, entre le fait d'accorder de la valeur [*valuing*] à quelque chose et le fait d'adopter une certaine forme d'attitude positive à son égard, laquelle fournit alors à quelqu'un une raison instrumentale d'agir. Nous ne croirions tout simplement pas quelqu'un qui affirmerait accorder une importance à l'honnêteté et qui, pourtant, ne montrerait jamais la moindre envie d'agir honnêtement lorsqu'il en aurait facilement l'occasion. Or il s'agit là d'un fait concernant la relation entre les valeurs *épousées* par un individu et ses raisons d'agir, et non pas d'un fait témoignant d'une relation entre l'évaluation morale et la motivation rationnelle.

Supposons, par exemple, que nous acceptions la caractérisation de Hume selon laquelle la justice est une vertu artificielle, dirigée vers le bien-être [*welfare*] général. Cela est, en un sens manifeste, une notion évaluative ou normative – « une valeur » au sens large tel qu'elle est employée dans le cadre de tels débats – et pourtant, il ne suit certainement pas de cette définition que tout être rationnel qui croit qu'un acte quelconque est juste en ce sens, et ce, peu importe ses désirs, aurait une raison instrumentale d'accomplir cet acte. Un individu rationnel peut ne pas accorder de valeur [*valuing*] à la justice pour elle-même [*for its own sake*], et peut avoir des fins contraires à celle-ci. Lorsque Hume discute de notre « obligation intéressée » à être juste, il semble reconnaître que, en fin de compte, il peut ne pas être possible de démontrer qu'un « fripon sensé » [*sensible knave*] a des raisons d'être juste. Bien entendu, Hume fait valoir que le reste d'entre nous – dont les cœurs se rebellent contre l'attitude du fripon rusé, lequel considère qu'il est acceptable de manquer à sa parole, de tricher ou de voler lorsque cela lui convient – a des raisons d'être juste, de considérer l'attitude du fripon comme injuste, et de se protéger contre ses prédatations.⁷

Le fripon pourrait affirmer, pour sa part, peut-être parce qu'il accepte l'analyse de la justice de Hume : « Oui, mon attitude est injuste ». Sur la base de la conception humienne de la relation entre la raison et les passions, le même fripon pourrait ajouter : « Mais qu'est-ce que cela peut me faire? », et ce, bien qu'il puisse être à même de saisir le contenu de son assertion précédente. Supposons que le fripon n'ait aucun doute quant à l'intelligibilité ou à la réalité du « bien-être général », et qu'il pense qu'il est parfaitement compréhensible que les individus attachent une grande importance à la notion de justice qui lui est corrélée dans le cadre de la vie en commun. Il réalise également que pour la majorité de l'humanité, dont les passions diffèrent des siennes, le fait d'être juste est une source et une condition de ce qui a le plus de valeur [*worthwhile*] dans la vie. Il comprend ainsi que les appels à la justice ont généralement une force motivante. En outre, il a lui-même recours à la catégorie de justice afin d'analyser le monde social, et il reconnaît – car ses calculs de fripon, en effet, en tiennent compte – la distinction entre, d'une part, les individus et les institutions qui sont véritablement justes et, de l'autre, ceux qui sont justes simplement en apparence ou sont communément considérés comme justes. Le fripon considère un bon nombre de concepts qui sont monnaie courante [*wide currency*] – les concepts religieux, par exemple – comme de simples fictions qui s'en prennent aux esprits

faibles, mais il ne considère pas la justice ainsi. Les esprits faibles et les moralistes, croit-il, ont couvert la justice de certains mythes – que la justice est sa propre récompense, que lorsque l'on voit ce qui est juste on aura automatiquement des raisons de faire ce qui est juste, et ainsi de suite. Mais cela étant dit, il croit que les esprits faibles et les moralistes ont également couvert la richesse et le pouvoir de certains mythes – que les riches ne sont pas véritablement heureux, que les puissants se précipitent vers leur chute, et ainsi de suite – et il ne doute pas, de ce point de vue, qu'il y ait des choses telles que la richesse et le pouvoir. Le fripon est heureux d'être affranchi des mythes prédominants concernant la richesse, le pouvoir et la justice; il est également heureux d'être libéré du fait de prêter autant ou peu d'attention, dans son esprit, à ces attributs, comme le commandent ses désirs ou ses circonstances. Il pourrait, par exemple, être convaincu du conseil de Mae West : il vaut tout à fait la peine d'acquérir des diamants, et « le bien [*goodness*] n'a rien à voir avec ça. »

Nous devons donc distinguer le travail consistant à déterminer ce à quoi un individu accorde de la valeur de celui visant à affirmer ce en quoi consiste pour un individu de prendre des mesures [*measurements*] à la lumière des critères d'une espèce d'évaluation qu'il reconnaît comme légitime.⁸

En niant la thèse de Hume concernant le caractère pratique du jugement moral et en supprimant par là même le fondement de la distinction humienne entre faits et valeurs, on ne nie pas pour autant le fait que la moralité possède cette caractéristique consistant à guider nos actions. La moralité peut certainement demeurer prescriptive au sein d'un cadre instrumental et peut se recommander à nous de la même façon dont, par exemple, l'épistémologie le peut : plusieurs fins humaines significatives et persistantes – bien que peut-être pas universelles – peuvent être poursuivies si nous appliquons certains critères d'évaluation à nos actions. Cela peut suffire à justifier à nous-mêmes notre souci persistant pour le statut épistémique ou moral de ce que nous faisons.⁹

En faisant valoir que la raison ne nous contraint pas à adopter des croyances ou des pratiques spécifiques, mises à part nos fins contingentes et variables, il peut sembler que je ne sois pas parvenu à composer avec le relativisme épistémique, et que j'aie miné l'argument en faveur du réalisme moral avant même de l'avoir présenté. La rationalité devient effectivement relative lorsqu'elle est instrumentale, mais cela n'est pas également le cas de l'épistémologie. La justification épistémique [*epistemic warrant*] de la croyance d'un individu peut être dissociée de la rationalité impliquée dans le fait d'entretenir cette croyance, car la justification épistémique peut être liée à un critère externe – comme c'est le cas, par exemple, dans le cadre des théories de la connaissance causales ou fiabilistes [*reliabilist*].¹⁰ Nous ne devons pas confondre l'idée selon laquelle nos fins en général sont bien servies par des mécanismes causaux et fiables de formation des croyances avec la thèse internaliste selon laquelle la raison nous contraint à adopter des moyens de ce type. Aux mécanismes fiables correspondent des coûts tout comme des bénéfices, et la poursuite efficace de certaines fins – le fripon ferait ici allusion à certaines fins religieuses, et à celles de certains moralistes – peut,

à certains égards, être incompatible avec l'adoption de moyens fiables d'investigation [*reliable means of inquiry*].

Cette réfutation de l'accusation de relativisme invite les partisans de la distinction fait-valeur à se déplacer sur un terrain ontologique. Peut-être que les faits et les valeurs ne peuvent représenter les pôles opposés d'une division épistémologique, marquée par ce que la raison et l'expérience peuvent nous interpeller à accepter. Néanmoins, l'idée de mécanismes causaux et fiables à la base de l'apprentissage moral et de faits moraux « dans le monde » sur lequel ces mécanismes agissent est sans doute si bizarre qu'il semble que je n'aie peut-être que multiplié mes difficultés.

III. LE RÉALISME DES VALEURS

Certes, l'idée d'une interaction causale avec la réalité morale serait intolérablement étrange si les faits moraux étaient considérés comme *sui genesis*.¹¹ Mais l'idée qu'il y a des mécanismes causaux à la base de l'apprentissage des faits moraux n'a rien de forcément bizarre si de tels faits sont constitués par des faits naturels. Telle est précisément la position que j'examinerai à présent. La réponse demeurera, toutefois, peu convaincante jusqu'à ce que l'on présente certains arguments positifs en faveur du réalisme à propos des faits moraux. À présent, tournons ainsi notre attention sur cette tâche.

Ce que l'on pourrait appeler « le stratagème générique du réalisme naturaliste » consiste à postuler un domaine de faits sur la base de la contribution qu'ils pourraient apporter à l'explication à posteriori de certains traits caractéristiques de notre expérience. Par exemple, un monde externe est postulé [*posited*] afin d'expliquer la cohérence, la stabilité et l'intersubjectivité de l'expérience sensible [*sense-experience*]. Un réaliste moral qui tire profit de cette stratégie doit montrer que la postulation de faits moraux peut, de manière similaire, avoir une fonction explicative. La stratégie peut réussir dans les deux cas seulement si la réalité postulée possède ces deux caractéristiques :

L'indépendance : La réalité existe et possède certains traits déterminés indépendamment du fait que nous pensions qu'elle existe ou qu'elle possède ces traits, voire indépendamment du fait que nous ayons de bonnes raisons de croire en cette réalité;

La rétroaction [feedback] : La réalité est constituée – et nous sommes constitués – de telle sorte que nous pouvons interagir avec elle, et ces interactions exercent de manière appropriée une influence et un contrôle constitutifs sur nos perceptions, nos pensées et nos actions.

Ces deux caractéristiques permettent au postulat [*posit*] réaliste de jouer un rôle dans l'explication de notre expérience, qui ne peut être remplacée sans perte par une simple *conception* de nous-mêmes ou de notre monde. Car bien que la médiation opérée par notre cadre conceptuel s'effectue même au niveau de nos

expériences perceptuelles les plus élémentaires, une réalité qui transcende l'expérience peut de multiples manières se faire sentir [*making itself felt*] causalement sans la permission de notre cadre conceptuel. Le succès ou l'échec de nos plans et de nos projets ne sont pas déterminés seulement par nos attentes. En résistant ou en cédant à des efforts mondains par des moyens qui ne sont pas anticipés par notre cadre conceptuel actuel, une réalité externe qui ne se révèle pourtant jamais directement dans la perception peut néanmoins influencer considérablement l'évolution subséquente de ce schème.

Le fait que le réalisme ait recours à un monde externe afin d'expliquer l'expérience sensorielle a souvent été critiqué sous prétexte qu'il ne s'agit là que d'une simple image. Mais avons-nous même une simple idée de ce à quoi peut ressembler une explication réaliste dans le cas des valeurs?¹² Je tenterai d'en esquisser une en présentant d'abord une conception de la valeur non morale : l'idée suivant laquelle quelque chose est désirable pour quelqu'un ou est bon [*good*] pour cet individu.¹³

Considérons d'abord l'idée d'*intérêt subjectif* propre à quelqu'un – ses désirs [*wants or desires*], conscients ou inconscients. Un intérêt subjectif peut être considéré comme une qualité secondaire, telle que le goût. Que je m'intéresse à quelque chose [*take a subjective interest in something*] revient à dire que cette chose a une *valeur* [*valence*] positive pour moi, c'est-à-dire que, dans des circonstances ordinaires, elle excite une attitude ou une inclination positive (pas nécessairement consciente) en moi. De la même manière, que j'affirme que le sucre me semble sucré revient à dire que, dans des circonstances normales, le sucre excite une certaine sensation gustative en moi. En tant que qualités secondaires, l'intérêt subjectif et la sucrosité éprouvée surviennent sur les qualités premières de celui qui perçoit, de l'objet (ou autre phénomène) perçu et du contexte environnant : celle ou celui qui perçoit est constitué de telle sorte qu'un tel objet dans un tel contexte excitera une sensation de ce type. Nous nommons la *base réductrice* de la qualité secondaire cet ensemble complexe de qualités premières, relationnelles et dispositionnelles.

Dans cette base réductrice s'offre à nous une notion objective qui correspond aux intérêts subjectifs et nous permet de les expliquer. Mais elle ne peut pas servir de fondement plausible à la notion de bien [*goodness*] non moral, étant donné que les intérêts subjectifs qu'elles fondent ne disposent pas d'une force normative suffisante pour rendre compte de l'idée de désirabilité [*desirableness*]. Les intérêts subjectifs sont souvent le reflet de l'ignorance, de la confusion ou du manque de considération, comme on peut le constater avec un certain recul. Le fait que je sois actuellement ainsi constitué que je désire quelque chose que, si j'en avais une meilleure connaissance, je souhaiterais n'avoir jamais désiré ne semble pas se présenter à moi comme ce qui fait partie de mon bien [*good*].

Pour remédier à ce défaut, permettons-nous d'introduire la notion d'*intérêt subjectif objectifié* d'un individu *A*, comme suit.¹⁴ Accordons à un individu *A* actuel des pouvoirs cognitifs et imaginatifs inconditionnels, de même qu'une

pleine connaissance de toute information factuelle et nomologique concernant sa constitution physique et psychologique, ses capacités, ses circonstances, son histoire, etc. A deviendra alors A^+ , lequel possède une connaissance complète et vive [*vivid*] de lui-même et de son environnement, et dispose d'une rationalité instrumentale qui ne peut en aucun cas faire défaut. Nous demandons ensuite à A^+ de nous dire non pas ce qu'*il* désire actuellement, mais ce qu'il voudrait que sa version non idéalisée de lui-même, A , veuille – ou, plus généralement, convoite – s'il devait se trouver lui-même dans la condition et les circonstances actuelles de A .¹⁵ Tout comme nous avons présupposé qu'il y a une base réductive au fondement des intérêts subjectifs actuels d'un individu A , nous présupposons qu'il y a une base réductive au fondement de ses intérêts subjectifs objectifiés, à savoir ces faits à propos de A et de ses circonstances que A^+ combinerait avec ses connaissances générales afin d'en arriver à des idées au sujet de ce qu'il voudrait vouloir s'il devait se trouver à la place de A .

Prenons, par exemple, Lonnie, un voyageur qui se trouve dans un pays étranger et qui se sent mal en point. Il désire avant tout venir à bout de son malaise et soulager son estomac, et il constate qu'il a un désir insatiable pour un goût familier : un grand verre de lait. Le lait est désiré par Lonnie, mais s'agit-il également de quelque chose de désirable pour lui? Lonnie-Plus peut voir que ce qui ne va pas avec Lonnie, outre le mal du pays, est la déshydratation, un mal qu'ont en commun plusieurs touristes, mais qui est rarement détectable au moyen du témoignage de l'introspection [*introspective evidence*]. L'effet de boire du lait difficilement digérable pourrait troubler davantage l'estomac de Lonnie et aggraver sa déshydratation. En revanche, Lonnie-Plus peut voir que l'ingestion d'une abondance de liquides clairs pourrait rapidement améliorer la condition physique de Lonnie – ce qui, incidemment, pourrait également l'aider en ce qui concerne son mal du pays. Lonnie-Plus peut également voir à quel point le fait de boire des liquides clairs serait désagréable pour Lonnie; ce qui arriverait précisément si Lonnie continuait à souffrir de déshydratation, et ainsi de suite. Sur la base de ces informations, Lonnie-Plus pourrait alors en venir à désirer tout cela s'il devait prendre la place de Lonnie; il voudrait boire des liquides clairs plutôt que du lait, ou du moins, agir de telle façon qu'un désir de ce type soit satisfait. La base réductive [*reduction basis*] de cet intérêt objectifié implique des faits au sujet de la constitution et des circonstances de Lonnie, lesquels déterminent, entre autres, ses goûts actuels et sa capacité de développer certains nouveaux goûts, les conséquences occasionnées par la déshydratation prolongée, les effets et la disponibilité de plusieurs types de liquides, et ainsi de suite.

Affirmons, par conséquent, que cette base réductive constitue la constellation des qualités premières qui fait en sorte que Lonnie a un certain *intérêt objectif*.¹⁶ C'est pourquoi nous pourrions affirmer que Lonnie a un intérêt objectif de boire des liquides clairs en vertu de cet ensemble de faits complexes, relationnels et dispositionnels. Autrement dit, nous pourrions affirmer que la base réductive, et non pas le simple fait que Lonnie-Plus aurait certains désirs [*wants*], est le vérificateur [*truth-maker*] de l'idée qu'il s'agit là d'un intérêt objectif de Lonnie. L'intérêt objectif explique ainsi la raison pour laquelle il existe un certain intérêt objectifié, et non pas l'inverse.¹⁷

Affirmons à présent que X est *non moralement bon* [good] pour A si et seulement si X satisfaisait un intérêt objectif de A .¹⁸ Nous pourrions penser au sujet des croyances de A^+ concernant ce qu'il voudrait vouloir s'il était à la place de A , qu'elles pourraient donner lieu à un classement des intérêts objectifs possibles de A , un classement qui pourrait refléter ce qui est le mieux ou le pire pour A , et qui nous permettrait de parler des désirs [wants] actuels de A en tant qu'il s'agit de meilleures ou de pires approximations de ce qui est le mieux pour lui. Nous pourrions aussi décomposer le point de vue de A^+ en des intérêts objectifs *prima facie* par opposition à des intérêts objectifs « tout bien considéré » [on balance] de A , où ces derniers se rapportent à la notion d'« un bien pour A » [a good for A] alors que les premiers renvoient à la notion « du bien de A » [the good for A].¹⁹

Il me semble qu'il s'agit là d'une explication plausible de ce en quoi consiste le bien [good] non moral de quelqu'un : grosso modo, ce qu'il voudrait lui-même convoiter s'il savait ce qu'il fait.²⁰

Qui plus est, cette explication conserve ce qui, selon moi, semble être une relation adéquate entre les valeurs non morales et la motivation. Supposons que l'on désire X , mais que nous nous demandions si X fait réellement partie de son bien [good]. Ce sentiment de perplexité survient généralement parce que l'on croit que l'on ne sait pas assez de choses sur X , sur soi ou sur son monde, ou bien parce qu'on sent que l'on n'est pas adéquatement rationnel ou réflexif [reflective] en évaluant l'information dont on dispose – on soupçonne peut-être que l'on a été captivé par quelques caractéristiques saillantes de X (ou bien repoussés par quelques caractéristiques saillantes d'options autres que X). Si l'on devait apprendre que l'on continuerait néanmoins à vouloir pour soi-même X dans des circonstances où l'on pourrait considérer les choses de manière à ce que l'on soit pleinement informé et rationnel, cela donnerait vraisemblablement moins de poids à l'inquiétude initiale. En revanche, si l'on devait apprendre que, étant pleinement informé et rationnel, on *ne* voudrait *pas* vouloir pour soi-même X dans ces circonstances, cela donnerait vraisemblablement plus de poids à cette inquiétude. Les désirs étant ce qu'ils sont, une inquiétude renforcée pourrait ne pas suffire pour éliminer le désir de X . Mais si l'on devait être sincèrement et vivement convaincu que le fait de désirer X n'est, en ce sens, pas appuyé par une réflexion exhaustive sur les faits, on pourrait vraisemblablement ressentir qu'il s'agit là d'une raison pour ne pas agir conformément à ce désir. L'ajustement du désir à la croyance ne saurait, dans un cas donné, être requis par la raison ou la logique; cet ajustement pourrait être « tout simplement [merely] psychologique ». Mais ce sont précisément des phénomènes psychologiques de ce type que les théories naturalistes de la valeur considèrent comme fondamentaux [basic].

Dans ce qui suit, nous devons avoir recours à la notion de « bien intrinsèque » [intrinsic goodness]. Nous affirmerons alors que X est un bien intrinsèquement non moral [intrinsically non-morally good] de A seulement si X constitue un intérêt objectif de A et n'implique aucune référence à tout autre intérêt objectif de A . Nous pouvons employer en un sens évident la notion d'intérêt intrinsèque

objectif afin de rendre compte de tout autre intérêt objectif. Étant donné que les individus et les environnements diffèrent à bien des égards, il n'est pas nécessaire de présupposer qu'ils ont tous les mêmes intérêts intrinsèques objectifs. À fortiori, nous n'avons pas besoin de présupposer qu'ils sont caractérisés par les mêmes intérêts instrumentaux objectifs. Nous devrions, toutefois, nous attendre à ce qu'il existe, à cet égard, des similarités correspondantes dans leurs intérêts lorsqu'il existe des similarités personnelles et situationnelles entre les individus, c'est-à-dire lorsqu'il existe des similarités dans les bases réductives.

Il est désormais possible de voir la façon dont la notion de bien [*goodness*] non moral se prête à des usages explicatifs. Pour commencer, elle peut nous permettre d'expliquer pourquoi les désirs actuels de quelqu'un peuvent avoir certaines caractéristiques contrefactuelles – par exemple, pourquoi quelqu'un aurait certains désirs hypothétiques plutôt que d'autres s'il devait être pleinement informé et conscient. Un usage explicatif de ce type – qui découle directement de la définition d'intérêt objectif – pourrait toutefois bien être jugé décevant à moins que d'autres fonctions explicatives soient trouvées.

Considérons alors la différence entre Lonnie et Tad, un autre voyageur se trouvant dans les mêmes lieux, mais qui, à la différence de Lonnie, désire boire des liquides clairs et agit conformément à ce désir. Tad sera ravivé alors que Lonnie demeurera léthargique. Nous pourrions expliquer cette différence en soulignant que, bien que Lonnie et Tad aient tous les deux agi en fonction de leurs désirs [*wants*], les désirs [*wants*] de Tad reflètent mieux ses intérêts. La congruence entre les désirs [*wants*] de Tad et ses intérêts peut être fortuite, tout comme il est possible que Tad soit au courant de sa déshydratation et connaisse le traitement standard à suivre. Dans ce dernier cas, nous affirmerions en général que l'explication de la différence de leurs conditions tient au fait que Tad, et non Lonnie, « savait ce qui était bien [*good*] pour lui. »

Nous pouvons anticiper que, en général, ce que A^+ voudrait vouloir s'il était à la place de A corrèlerait bien avec ce qui permettrait à A d'éprouver un état de bien-être physique ou psychologique, et d'échapper à un état de malaise physique ou psychologique. Bien sûr, notre état de bien-être ou de malaise figure parmi les choses qui nous importent le plus, et ce, de manière la plus fiable, même après mûre réflexion.²¹ Le recours à des degrés de congruence entre les désirs [*wants*] et les intérêts de A nous aidera souvent à expliquer les faits à propos du degré de satisfaction de A concernant sa propre vie. L'explication ne pourrait être préservée si l'on substituait « le fait de croire qu'il y a congruence » avec « le fait d'être congruent (à un tel ou tel degré) », étant donné que, comme le démontrent des cas comme celui de Lonnie, même si nous parvenions à nous convaincre nous-mêmes que nos désirs [*wants*] reflètent nos intérêts, le fait d'agir conformément à ces désirs [*wants*] pourrait ne pas engendrer beaucoup de satisfaction.

En vertu de la corrélation escomptée entre le fait d'agir en fonction de motifs qui concordent avec nos propres intérêts et le fait de parvenir à un certain niveau de

satisfaction et d'éviter la détresse, nos intérêts objectifs peuvent aussi jouer un rôle explicatif dans l'évolution de nos désirs. Considérons ce que j'appelle le *mécanisme désirs* [wants]-*intérêts*, lequel permet aux individus d'apprendre, sur la base de leurs expériences, ce en quoi consistent leurs intérêts dans un état à la fois de conscience-de-soi [*selfconscious*] et de non-conscience-de-soi [*unself-conscious*]. Dans les cas les plus simples, l'apprentissage par essais et erreurs conduira à une rétention sélective de désirs [wants] qu'il est possible de satisfaire [*satisfiable*] et à des résultats satisfaisants pour l'agent.

Supposons, par exemple, que Lonnie cède à son envie de boire le lait. Peu de temps après, son état se dégrade. Étant toujours incapable d'identifier la source de son malaise et resté aux prises avec le désir de ce qui est familier, son attention se tourne vers un signe vert et rouge dans la fenêtre d'un petit commerce sur son passage, indiquant « 7-Up ». Il se précipite à l'intérieur et achète une bouteille. Bien que tiède, il la boit avec impatience. « Mmm », pense-t-il, « j'en prendrais une autre ». Il achète une seconde bouteille et la boit. À ce moment, sa soif de soda tiède est éteinte et il poursuit son chemin. Après quelques heures, son humeur s'améliore. Lorsqu'il passe de nouveau devant le commerce sur le chemin vers son hôtel, l'association plaisante liée au fait de boire du 7-Up l'incite à en acheter de nouveau et à l'emporter avec lui. Ce soir-là, dans la sombre solitude de sa chambre, il retrouve le goût familier et rassurant du soda réconfortant, et boit quelques bouteilles de plus avant de finalement s'endormir. Au moment de son réveil le lendemain matin, il se sent beaucoup mieux. Pour résumer une histoire ennuyeuse : la prochaine fois que Lonnie se sentira mal en point à l'étranger, il pourrait avoir une certaine tendance consciente ou inconsciente, raisonnée ou superstitieuse d'être à la recherche du 7-Up. S'il n'est pas en mesure d'en trouver, il pourrait alors se mettre à la recherche d'un soda à saveur de lime, spécifique au pays où il se trouve, ou peut-être même l'*agua mineral con gaz* qu'il méprisait autrefois. Peu à peu, au fur et à mesure que Lonnie voyage davantage et souffre d'un malaise semblable, il en viendra à boire régulièrement des liquides clairs et, de manière régulière, à se sentir mieux, de sorte qu'il développera finalement un désir actuel pour de tels liquides – et une aversion pour d'autres liquides, tels que le lait – dans de telles circonstances.

C'est ainsi que les désirs de Lonnie ont évolué au cours de son expérience pour se conformer de plus près à ce qui est bien pour lui, au sens naturaliste du terme. Le processus ne relève pas d'une réponse idéalement rationnelle au fait de recevoir des informations idéales, mais plutôt d'une expérimentation en grande partie non réfléchie, accompagnée d'associations et de renforcements positifs et négatifs. Il n'y a aucune garantie que les désirs « appris » par de la rétroaction de ce type reflètent précisément ou complètement le bien [*good*] de l'individu. Il est encore moins certain que, même lorsqu'un ajustement adéquat dans les désirs se produit, l'agent comprenne l'origine de ses désirs ou soit capable de se représenter la nature des intérêts qu'ils reflètent. Mais il s'agit alors d'une caractéristique tout à fait générale propre aux divers moyens par lesquels nous apprenons à connaître le monde, qu'ils puissent ne pas fournir des représentations précises [*accurate*] ou complètes de ce monde. Ma capacité de percevoir

et de comprendre mon environnement coexiste avec, voire s'appuie sur les mêmes mécanismes que ma disposition [*liability*] à être sujet à la duperie par la voie d'illusions, d'attentes ou d'apparences.

Il existe des raisons largement théoriques de penser qu'il existe quelque chose comme un mécanisme désirs [*wants*]-intérêts et qu'il a un rôle important dans la formation de désirs. Les êtres humains sont des créatures principalement motivées par des désirs plutôt que par des instincts. Si de telles créatures étaient incapables au cours de leurs expériences d'adapter leurs désirs [*wants*] de manière plus étroite à leurs intérêts essentiels – peut-être parce qu'elles ne sont pas plus susceptibles de faire l'expérience d'états internes positifs lorsque leurs intérêts essentiels sont servis plutôt que lorsqu'ils ne le sont pas –, nous serions portés à croire qu'elles ne connaîtraient pas un long et fructueux avenir. Ainsi, si les êtres humains en général n'en étaient pas venus à manger le type de nourriture qui leur assure un certain degré de bien-être physique, ou à s'engager dans le type d'activités ou de relations qui leur permet de maintenir une bonne santé mentale, nous ne serions pas ici aujourd'hui en train de nous interroger sur ce qui est bon [*good*] pour nous. Étant donné que des créatures aussi sophistiquées et complexes que les êtres humains ont évolué dans des contextes liés à une diversité d'environnements et ont, par ailleurs, pris l'habitude de modifier leurs environnements, nous devrions alors nous attendre à ce que nous fassions preuve d'une flexibilité considérable quant à notre capacité d'adapter, au cours de notre expérience, nos désirs [*wants*] à nos intérêts. Toutefois, cette même flexibilité fait en sorte qu'il s'agit là d'un mécanisme qui n'est pas fiable [*unreliable*] : nos désirs [*wants*] peuvent, à n'importe quel moment, considérablement différer de nos intérêts, et ce, de manière arbitraire. En outre, nous pourrions ne pas être en mesure d'avoir des expériences qui pourraient nous amener à remarquer ceci, ou de faire suffisamment l'objet de rétroaction pour avoir la chance de développer de nouveaux désirs [*wants*] qui se rapprochent davantage de nos intérêts. Il est tout à fait possible, et rarement peu fréquent, qu'un individu vive le reste de sa vie normale sans jamais reconnaître certains de ses intérêts les plus fondamentaux ou s'ajuster à eux. Les mécanismes culturels d'acquisition de désirs [*wants*] peuvent en partie remédier aux limitations individuelles, et favoriser l'apprentissage et même la théorisation sur l'étendue de plusieurs vies et durées de vie, mais ces mêmes mécanismes créent aussi un vaste potentiel d'inculcation de désirs [*wants*] qui vont à l'encontre d'intérêts.

On peut conférer à la thèse selon laquelle il existe un mécanisme désirs [*wants*]-intérêts relativement le même statut et la même plausibilité décontractée [*breezy*] que l'argument plus strictement biologique selon lequel nous devons nous attendre à ce que l'œil humain soit capable de détecter des objets d'une taille et d'une forme similaires à celles de nos prédateurs ou de nos proies. Il n'est pas nécessaire de ne supposer rien de tel qui correspond plus ou moins à l'infailibilité. Il nous suffit de présupposer assez de succès fonctionnel, en vertu duquel nous parvenons à tenir notre rang dans un monde très souvent inhospitalier.²²

Jusqu'ici, l'argumentaire n'a porté que sur les intérêts objectifs qui peuvent être caractérisés en tant que besoins, mais le mécanisme désirs-intérêts peut s'appliquer à n'importe quel intérêt – même les intérêts liés aux aptitudes particulières ou aux rôles sociaux d'un quelconque individu – dont la frustration s'accompagne – et ce, même indirectement – de résultats consciemment ou inconsciemment insatisfaisants pour cet individu. (Bien entendu, plus l'association est indirecte, plus il est improbable que le mécanisme soit fiable.) Par exemple, l'expérience consistant à suivre des cours de mathématiques et de philosophie pourrait inciter une étudiante qui se croyait jusque-là destinée à une carrière de mathématicienne à en venir à préférer une carrière de philosophe, laquelle serait en fait mieux adaptée à ses aptitudes et à ses attitudes. Un ouvrier récemment promu à un poste de gestionnaire pourrait se trouver moins enclin à répondre aux griefs des employés – ce qu'il revendiquait pourtant auparavant auprès de ses gestionnaires –, alors que ses anciens collègues pourraient, de même, se trouver moins enclins qu'auparavant à se confier à lui.

Si un mécanisme désirs-intérêts est postulé, et si ce qui est non moralement bien pour un individu est une question qui relève de son intérêt objectif, nous pouvons alors affirmer que les valeurs objectives sont à même de jouer un rôle dans l'explication des valeurs subjectives du type requis par le réalisme naturaliste des valeurs. Ces explications fournissent même un appui à certaines prédictions qualifiées [*qualified predictions*] : par exemple, que, toutes choses étant égales par ailleurs, les individus seront habituellement les mieux placés que de tierces parties pour juger de leurs propres intérêts; que des individus possédant des caractéristiques personnelles et sociales similaires tendront à avoir des valeurs similaires; et qu'un consensus général plus important pourrait être dégagé sur ce qui est désirable dans les domaines de la vie où les individus sont semblables à d'autres égards (par exemple, sur le plan des motivations de base) et où l'on prévoit que les mécanismes d'essais et erreurs fonctionnent bien (par exemple, lorsqu'un savoir ésotérique n'est pas requis). Je ne suis pas en mesure de me prononcer sur l'exactitude de ces prédictions, mais celles-ci concordent avec des opinions largement répandues, ce qui est tout à son honneur.

Il conviendrait peut-être de souligner que, bien que je parle d'objectivité des valeurs, les valeurs en question sont humaines et existent seulement parce que les êtres humains existent. Du point de vue d'une théorie de la valeur classique [*old-fashioned*], il s'agit d'une notion relationnelle – plutôt qu'absolue – de ce qui est bon [*goodness*]. Bien qu'ils soient relationnels, les faits pertinents au sujet des êtres humains et de leur monde sont objectifs de la même façon dont les entités non relationnelles, telles que les pierres, le sont : leur existence ou leur nature ne dépend pas simplement de la conception que nous en avons.²³

Compris ainsi, les intérêts objectifs surviennent sur les faits naturels et sociaux. Est-ce que cela signifie que les intérêts objectifs ne peuvent avoir aucune valeur explicative, étant donné qu'il devrait toujours être possible, en principe, de rendre compte de n'importe quel fait particulier qu'ils prétendent expliquer par référence seulement à la base de survenance? Si la survenance était à elle seule

une raison de refuser un rôle explicatif à une série de concepts quelconques, nous devrions alors affirmer que la chimie, la biologie et le génie électrique – lesquels surviennent manifestement sur la physique – sont dépourvus de pouvoir explicatif. En effet, même la possibilité d'une réduction catégorique n'est pas une raison de douter de ce pouvoir explicatif. Le fait d'établir une relation de réduction entre, par exemple, un phénomène chimique tel que la valence et un modèle physique de l'atome ne suggère aucunement qu'il n'y a pas de phénomène tel que la valence, ou que les généralisations se référant à la valence ne peuvent fournir un appui à des explications. Il ne peut être question ici d'économie ontologique ou de rejeter des entités non nécessaires, tel que ce serait le cas si, par exemple, la valence était effectivement considérée comme un phénomène *sui generis*, au-delà de toute constellation de propriétés physiques. Les faits décrits au moyen des principes de la valence chimique sont authentiques, et ils permettent d'établir une systématisation puissante et explicative des combinaisons chimiques. Qu'une réduction de la chimie à la physique atomique ait été accomplie ne fait que renforcer ces affirmations.

Nous sommes persuadés que la notion de valence chimique est explicative puisque les explications formulées en termes de valence chimique inscrivent des *explananda* dans un cadre nomique distinct et bien articulé, de manière à ce qu'elle en augmente notre compréhension de manière évidente. Mais existe-t-il une théorie aussi puissante et explicative de la notion d'intérêt objectif, qui peut nous donner des raisons de penser – peu importe si une réduction stricte est possible ou non – que les explications élaborées au moyen de cette notion sont véritablement informatives?

Le type de réalisme des valeurs esquissé ici m'apparaîtrait inintéressant s'il me semblait qu'aucune théorie d'importance ne pourrait être développée au moyen de la catégorie de « valeur objective ». Mais en décrivant le mécanisme désirs-intérêts, j'ai déjà tenté d'indiquer qu'une théorie de ce genre est possible. Lorsque nous cherchons à expliquer les raisons pour lesquelles ils agissent ainsi, les raisons pour lesquelles ces individus ont certaines valeurs ou certains désirs, et ce qui parfois les conduit au conflit et, à d'autres moments, à la coopération, il est tout fait naturel, du point de vue du sens commun de même que des sciences sociales, de parler en termes d'intérêts individuels. De telles explications demeureront pourtant incomplètes et superficielles tant et aussi longtemps que nous nous en tiendrons seulement aux intérêts subjectifs, puisque ceux-ci doivent également faire l'objet d'une explication.²⁴

IV. LE RÉALISME NORMATIF

Supposons que tout ce que j'ai affirmé jusqu'à maintenant est généreusement tenu pour acquis. Je n'aurais pourtant pas encore le droit de parler de réalisme *moral* puisque je n'aurais fait valoir que la possibilité d'un type de réalisme à propos de ce qui est non moralement bon [*non-moral goodness*], une notion pouvant même être admise par de parfaits sceptiques en matière de moralité. Afin de pouvoir parler de réalisme moral, je dois montrer comment le réalisme

des valeurs typiquement morales ou des normes morales est possible. Je me concentrerai sur les normes morales, c'est-à-dire sur ce qui a trait au bien et au mal moral [*matters of moral rightness and wrongness*]; mais l'argument que je présente peut, par extension, s'appliquer également aux valeurs morales. Ce choix s'explique, en partie, par le fait que le réalisme normatif semble intuitivement beaucoup moins plausible que le réalisme des valeurs. Il n'est donc pas surprenant que plusieurs propositions actuelles mises de l'avant en faveur du réalisme moral se concentrent essentiellement sur la question des valeurs – et parfois seulement sur ce qui constitue des valeurs non morales. Toutefois, peu importe la conception de la moralité proposée, une théorie morale doit présenter une conception du bien [*rightness*].

Le réalisme moral normatif n'est pas plausible pour diverses raisons, mais dans le cadre du présent essai, la raison plus pertinente tient au fait qu'il me semble impossible d'étendre la stratégie générique du réalisme naturaliste aux normes morales. Quelle place revient aux faits concernant ce qui *doit* être le cas dans le cadre d'une explication? Est-ce que les faits concernant la façon dont les choses *sont* dans le monde [*the way things are*] suffisent pour expliquer la façon dont on doit agir? Bien sûr qu'il en est ainsi. Mais alors, ma conception naturaliste du réalisme moral m'engage à considérer les faits au sujet de ce qui doit être le cas comme une espèce particulière de faits concernant la façon dont les choses sont dans le monde. Par conséquent, il est possible que ces faits puissent avoir une fonction dans le cadre d'une théorie explicative. Pour envisager comment cela peut être possible, je propose que l'on considère quelques exemples d'explications impliquant des normes naturalisées en dehors du domaine de la moralité.

« Pourquoi le toit s'est-il effondré? – Pour qu'une maison puisse supporter des charges de neige de cette teneur, la charpente doit être constituée de planches de bois qui mesurent au moins 2 pouces x 8 pouces, et non pas 2 pouces x 6 pouces ». Cette explication est tout à fait acceptable en soi, mais elle implique un « doit » [*ought*]. Nous pourrions, bien sûr, éliminer ce « doit » [*ought*] comme suit : « s'il est pour être à même de résister à de telles chutes de neige, un toit conçu en fonction de ce modèle est alors dans l'obligation de disposer d'une charpente avec une coupe transversale d'au moins 2 pouces x 8 pouces ». Au « doit » [*ought*] architectural se substitue un « si... alors... » qui relève de l'ingénierie. Cela est possible puisque le « doit » est clairement hypothétique, ce qui reflète l'objectif architectural, universel consistant à fabriquer des toits suffisamment forts pour qu'ils ne s'effondrent pas. Puisque l'objectif est fixé selon le contexte, qu'il y a des réponses plus ou moins définies apportées à la question de savoir comment cet objectif peut être réalisé et, en outre, que le phénomène à expliquer est le résultat d'un processus qui implique une sélection contre les instances qui n'atteignent pas cet objectif, l'explication qui implique le « doit » transmet de l'information explicative.²⁵ Je qualifierai ce type d'explication de « *critérial* » : nous expliquons pourquoi quelque chose se produit par référence à un critère pertinent, étant donné l'existence d'un processus qui sélectionne en faveur (ou contre) des phénomènes qui se rapprochent approximativement de ce critère. Bien que le critère soit défini sous l'angle naturaliste, il pourrait en

même temps être de nature à jouer un rôle régulateur dans le cadre des pratiques humaines; en l'occurrence, la construction de bâtiments.

Un type plus familier d'explication critériale concerne les normes de rationalité individuelle. Considérons le recours à une théorie instrumentale de la rationalité afin d'expliquer le comportement d'un individu à la lumière de ses croyances et de ses désirs, ou de rendre compte de la manière dont ses croyances se modifient avec l'expérience.²⁶ Bobby Shaftoe est parti en mer parce qu'il croyait que c'était la meilleure façon de faire fortune, et il voulait avant tout faire fortune. Son coéquipier Reuben Ramsøe en est venu à la conclusion qu'il n'était pas apprécié par les autres matelots puisqu'il constatait qu'ils se moquaient de lui et ne dissimulaient pas leur appréciation pour les flagellations constantes qui lui étaient infligées par le second capitaine. La validité de ces explications tient au fait que l'action ou la croyance en question était tout à fait rationnelle du point de vue de l'agent qui se trouvait dans ces circonstances, et parce que nous supposons correctement que Shaftoe et Ramsøe devaient être tout à fait rationnels.

Les faits portant sur les degrés de rationalité instrumentale s'inscrivent dans le cadre d'explications par d'autres moyens également. Considérons d'abord la question de savoir pourquoi Bobby Shaftoe a connu plus de succès que la plupart de ses semblables dans la réalisation de ses buts. Nous pourrions attribuer son succès au fait que Shaftoe est mieux doté de raison instrumentale que la plupart des individus – peut-être possède-t-il une perspicacité supérieure à la moyenne pour estimer la probabilité des conséquences [*outcomes*], ou bien fait-il preuve d'une fiabilité supérieure à la moyenne en matière d'inférences déductives, ou dispose-t-il d'une imagination supérieure à la moyenne pour évaluer diverses options.

En deuxième lieu, bien que nous soyons tous imparfaits en matière de délibération, notre comportement peut en venir à incarner des habitudes et des stratégies qui nous permettent de nous rapprocher d'une rationalité optimale de manière plus étroite que ce que nos défauts de délibération [*deliberative defects*] pourraient nous faire croire. Le mécanisme est simple. Des schémas [*patterns*] de croyances et de comportements qui manifestent une faible rationalité instrumentale tendent à être, dans une certaine mesure, voués à l'échec [*self-defeating*]; une incitation qui favorise leur modification. En revanche, les schémas qui manifestent une forte rationalité instrumentale tendent à être, dans une certaine mesure, gratifiants [*rewarding*]; une incitation qui favorise leur reproduction. Ces incitations peuvent influencer nos croyances et nos comportements même si les inconvénients ou les avantages des schémas en question ne font pas l'objet d'une délibération consciente. Dans de tels cas, nous pourrions affirmer que nous acquérons ces habitudes ou ces stratégies parce qu'ils *sont* plus rationnels, sans qu'il y ait une quelconque *croyance* intermédiaire qui donne à penser cela. Les psychologues cognitifs ont ainsi exposé certaines des stratégies ou heuristiques inconscientes que nous employons afin de permettre à nos intellects limités de trier davantage de données et de porter des jugements plus rapides et plus constants que ce serait le cas si nous avions seulement recours à

des formes plus standards de raisonnement explicite.²⁷ À notre insu, nous en venons à dépendre d'heuristiques en partie parce qu'elles sont sélectivement renforcées par leurs avantages instrumentaux par rapport au raisonnement standard et explicite, c'est-à-dire en raison de leur plus grande rationalité. De même, nous pouvons – sans que nous le réalisions ou nous puissions l'admettre à nous-mêmes – développer des schémas de comportements qui encouragent ou découragent des comportements particuliers chez autrui, tels que le mécanisme inconscient par lequel nous causons chez les individus dont nous n'apprécions pas la compagnie le fait de ne pas apprécier la nôtre. Finalement, en tant qu'enfants, nous étions pratiquement incapables de procéder à des évaluations rationnelles lorsque l'acquisition d'un gain éloigné dans le temps impliquait une perte imminente. Pourtant, d'une certaine manière, nous sommes parvenus de manière largement non délibérative, avec le temps, à acquérir diverses habitudes intéressantes, telles que la capacité de déplacer vers la périphérie de notre attention certaines pensées vives au sujet de l'avenir immédiat, capacité qui nous permet en tant qu'adultes de nous rendre par nous-mêmes chez le dentiste sans être poussés par derrière. L'explication critérielle en termes de rationalité individuelle s'étend ainsi aux comportements bien au-delà de l'action délibérée. Il devient alors possible de voir dans l'émergence de comportements de ce genre, tout comme dans le mécanisme désirs-intérêts, quelque chose que nous pourrions nommer sans déformation « apprentissage ».

En effet, notre tendance à développer au fur et à mesure des habitudes et des stratégies rationnelles peut coopérer avec le mécanisme désirs-intérêts afin de servir de base à une forme *étendue* d'explication critérielle, où la rationalité d'un individu est évaluée non pas en fonction de ses croyances et de ses désirs actuels [*occurrent*], mais en fonction de ses intérêts objectifs. Les exemples concernant le mécanisme désirs-intérêts que nous avons considérés antérieurement impliquent des éléments d'une explication de ce type, en ce qu'ils démontrent non seulement que les désirs sont ajustés en fonction des intérêts, mais aussi que le comportement est ajusté en fonction de désirs nouvellement ajustés. Sans la présence d'une modification adéquate du comportement, qui reflète les désirs en évolution, la rétroaction nécessaire à notre prise de connaissance de nos désirs ne se produirait pas. Avec une modification de ce type, le comportement peut lui aussi devenir plus rationnel au sens élargi du terme. Un individu qui fait preuve de rationalité instrumentale peut être disposé à ajuster ses moyens en fonction de ses fins, mais ce qui peut résulter du fait d'avoir recours à un moyen particulier – choisir un cours, accepter un nouvel emploi – peut être une évaluation plus informée de ses fins et peut-être une reconsidération de celles-ci.

La théorie de la rationalité individuelle – que ce soit dans sa forme simple ou étendue – nous offre ainsi une instance du type requis pour l'élaboration d'un exemple de réalisme normatif. Les évaluations des degrés de rationalité instrumentale jouent un rôle majeur dans nos explications du comportement des individus, mais elles fonctionnent aussi simultanément à titre de force normative pour l'agent. Peu importe les autres soucis que peut avoir un agent, il s'agit sûrement, pour elle ou lui, d'une caractéristique positive de l'action que l'efficacité

de celle-ci soit relative à ses croyances ou à ses désirs ou, dans un sens étendu, relative aux croyances et aux désirs qui reflètent adéquatement ses conditions et ses circonstances.

La force normative manifestée par ces théories de la rationalité individuelle ne doit pas, toutefois, être simplement dérivée de leur usage explicatif. On peut employer une théorie de la rationalité instrumentale afin d'expliquer des comportements et la rejeter à titre de théorie normative des raisons, tout comme on peut expliquer une action par son irrationalité sans pour autant alors promouvoir la déraison.²⁸ La relation entre les rôles normatifs et explicatifs de la conception instrumentale de la rationalité doit plutôt être reconduite à leur terrain commun : le système motivationnel humain. Il s'agit d'un fait, en ce qui nous concerne, que nous avons à la fois des fins et la capacité d'entreprendre des actions délibérées en fonction de celles-ci et d'ajuster non délibérément notre comportement relativement à nos fins. Par conséquent, des options s'offrent sur des chemins de divers horizons de possibilités que nous évaluons différemment. Lorsque nous expliquons les raisons à l'origine des choix des personnes et les causes de leur comportement, de même que lorsque nous faisons appel à leurs intuitions quant à la rationalité de leurs décisions ou de leurs actions éventuelles, nous œuvrons sur ce terrain puisque nous faisons appel dans la mesure du possible aux faits concernant ce qui motive les agents dans les faits ou en principe.

C'est ainsi qu'émerge la possibilité d'affirmer que des faits existent au sujet de ce que les individus ont des raisons de faire, des faits qui peuvent être considérablement indépendants de – et plus probants normativement que – l'idée que se fait actuellement l'agent de ses raisons. L'argument en faveur d'un tel réalisme à propos de la rationalité individuelle n'est pas plus probant que les arguments soutenant la double proposition selon laquelle la conception de la rationalité instrumentale et individuelle en question est caractérisée à la fois par un pouvoir explicatif et par un type de force approbative qu'une théorie des *raisons* se doit de posséder, mais ces arguments (bien que je n'en discuterai pas davantage ici) semblent, selon moi, très convaincants.

* * *

Procédons maintenant au-delà de la théorie de la rationalité individuelle et demandons-nous à quoi pourraient ressembler des explications critérielles [*criteria explanations*] qui impliquent des normes typiquement morales. Afin de répondre à cette question, nous devons savoir ce qui distingue une norme morale d'autres critères d'évaluation. L'évaluation morale semble porter avant tout sur l'évaluation de conduites ou de caractères où les intérêts de plus d'une personne sont en jeu. Qui plus est, l'évaluation morale évalue les actions ou les conséquences [*outcomes*] de manière singulière : les intérêts du plus fort ou du plus prestigieux ne prévalent pas toujours, des raisons purement prudentielles peuvent être subordonnées, et ainsi de suite. Plus généralement, les résolutions morales sont considérées comme déterminées par des critères de choix qui sont *non indexicaux* et, en un certain sens, *englobants* [*comprehensive*]. Cela a incité un

certain nombre de philosophes à vouloir rendre compte du caractère spécifique de l'évaluation morale en identifiant un *point de vue moral* qui est impartial, mais qui prend également en considération tous les individus pouvant être éventuellement concernés. D'autres théoriciens en éthique en sont venus à une conclusion similaire en examinant les types de raisons que nous considérons typiquement comme pertinents ou non pertinents dans le cadre du discours moral. Poursuivons sur ces voies. Nous pouvons ainsi affirmer que les normes morales reflètent un certain type de rationalité, une rationalité qui relève non pas du point de vue d'un individu en particulier, mais de ce que nous pourrions appeler un point de vue social.²⁹

En elle-même, l'adéquation [*equation*] entre le bien moral [*moral rightness*] et la rationalité d'un point de vue social n'est pas extrêmement restrictive, car selon ce que l'on entend par rationalité, cette adéquation pourrait être établie par un utilitariste, un kantien ou même un non cognitiviste. Il doit en être ainsi puisque si cette adéquation doit pouvoir saisir ce qui est spécifique aux normes morales, elle doit être compatible avec le plus vaste éventail de théories morales reconnues. Toutefois, dès que l'on opte pour une conception particulière de la rationalité – telle que la conception de la rationalité conçue comme la recherche efficiente du bien non moral [*the non-morally good*], ou comme autolégislation autonome et universelle, ou comme l'expression non cognitive d'une approbation hypothétique [*hypothetical endorsement*] – cette caractérisation schématique commence alors à assumer un contenu moral. Nous avons adopté ici la conception instrumentale de la rationalité, et cela – en lien avec la conception de ce qui est non moralement bon (*non-moral goodness*) – signifie que l'argument en faveur du réalisme moral présenté ci-dessus consiste en un argument qui présuppose et prétend défendre une théorie morale particulière et substantielle.³⁰

En quoi consiste cette théorie? Introduisons d'abord une idéalisation de la notion de rationalité sociale en considérant ce qui serait rationnellement approuvé si tous les intérêts de tous les individus éventuellement concernés étaient pris en considération dans des circonstances où l'on dispose d'informations vives et complètes.³¹ Étant donné la supposition suivant laquelle nous disposons d'informations vives et complètes, les intérêts en question seront objectifs. Considérant la théorie du bien [*goodness*] proposée dans la section III, cette idéalisation correspond à ce qui est rationnel d'un point de vue social, relativement à la réalisation du bien intrinsèque non moral [*intrinsic non-moral goodness*]. Il semble, selon moi, qu'il s'agit là d'un critère reconnaissable et plausible, voire très peu controversé du bien moral [*moral rightness*]. Le bien moral relatif [*relative moral rightness*] consiste en un certain degré d'approximation de ce critère.

La question se pose maintenant de savoir si la notion de degrés de bien moral [*moral rightness*] peut intervenir dans les explications des comportements ou des processus d'apprentissage moral qui existent parallèlement aux usages explicatifs que permet la notion de degrés de rationalité individuelle, comprise, en particulier, en un sens étendu. Je tenterai de suggérer divers moyens qui rendraient possibles de tels usages.

Tout comme l'individu qui néglige de manière considérable certains de ses intérêts sera susceptible [*liable*] d'éprouver diverses formes d'insatisfaction, de même une organisation sociale [*social arrangement*] – par exemple, une forme de production, une hiérarchie sociale ou politique, etc. – qui dévie par rapport à la rationalité sociale en négligeant considérablement les intérêts d'un groupe spécifique sera potentiellement sujette à l'insatisfaction et à l'agitation. Que ce potentiel soit réalisé dépend d'un ensemble considérable de circonstances. En raison de la socialisation, ou d'autres limitations s'exerçant sur l'expérience ou la connaissance des membres de ce groupe, le mécanisme désirs-intérêts ne pourrait pas avoir fonctionné de manière à ce que les désirs de ses membres reflètent leurs intérêts. Par conséquent, les individus peuvent n'éprouver aucune frustration directe de leurs désirs malgré la mise à l'écart de leurs intérêts. Le groupe peut également être trop dispersé ou trop faible pour se mobiliser de manière efficace. Il se peut également qu'il fasse l'objet d'une répression sévère. D'autre part, certaines circonstances sociales et historiques favorisent la réalisation de ce potentiel d'agitation sociale, telles que, par exemple, le fait de fournir aux membres de ce groupe des expériences qui ont pour conséquence de les rendre plus susceptibles de développer des intérêts s'arrimant à leurs désirs, d'affaiblir l'appareil de répression actuel, de fournir aux membres de ce groupe un accès inédit à des ressources ou à de nouvelles opportunités de mobilisation, ou simplement de dissiper l'illusion que le changement est impossible. On peut s'attendre à ce que le potentiel d'agitation sociale se manifeste alors dans de telles circonstances.

Tout comme les explications impliquant des évaluations de la rationalité individuelle ne peuvent être remplacées par des explications impliquant des *croyances individuelles à propos* de ce qui doit être rationnel, de même les explications impliquant des évaluations au sujet de la rationalité sociale ne peuvent être remplacées par des *croyances à propos* de ce qui est moralement bien [*right*]. Par exemple, un sentiment de mécontentement peut surgir lorsqu'une société dévie par rapport à la rationalité sociale, mais non pas en conséquence de la croyance que cela est effectivement le cas. Supposons que tous les membres d'une société quelconque croient que celle-ci est juste. Cette croyance peut contribuer à la stabilisation de cette société, mais si, dans les faits, les intérêts de certains groupes sont mis à l'écart, un potentiel d'agitation sociale se manifestera alors sous différentes formes (l'aliénation, une baisse du moral, un déclin de l'efficacité de l'autorité, et ainsi de suite) bien avant que des changements se produisent dans les croyances à propos de ce qui est juste [*justness*] relativement à cette société. Cela nous permettra, en outre, d'expliquer la raison pour laquelle les membres de certains groupes en viennent à croire que cette société est injuste, si ceux-ci croient effectivement que c'est le cas.

Outre le fait qu'elles sont caractérisées par un certain type de potentiel d'agitation sociale, les sociétés qui ne parviennent pas à se rapprocher de la rationalité sociale peuvent également partager des caractéristiques supplémentaires : elles peuvent manifester une tendance vers certaines doctrines religieuses ou idéologiques, ou vers certains types d'appareils de répression; elles peuvent être moins

productives à certains égards (par exemple, ne pas être en mesure de développer certaines ressources humaines) et plus productives dans d'autres contextes (par exemple, dans l'extraction à moindre coût d'une plus grande force de travail à partir de certains groupes), et peuvent donc être économiquement prospères de manière différentielle selon les conditions de production avec lesquelles elles doivent composer, etc.

Si une notion de rationalité sociale peut de manière légitime être partie intégrale des explications empiriques des phénomènes de ce type, il doit alors être possible de formuler une caractérisation informative des circonstances dans lesquelles les déviations par rapport à, ou les approximations de la rationalité sociale peuvent être considérées comme conduisant vraisemblablement à des conséquences [*outcomes*] sociales particulières; notamment, une caractérisation informative des conditions dans lesquelles des groupes dont les intérêts sont sacrifiés pourraient vraisemblablement manifester ou mobiliser une forme de mécontentement. Bien qu'on ne puisse pas savoir a priori si une explication de ce type est possible, on peut néanmoins constater dans le cadre des travaux récents en histoire sociale et en sociologie historique l'émergence des divers éléments d'une théorie visant à expliquer quand et comment un potentiel persistant de mécontentement social en vient à se manifester suite au fait que des intérêts persistent à être sacrifiés.³²

Un individu dont les désirs [*wants*] ne reflètent pas ses intérêts ou qui ne parvient pas à la rationalité instrumentale peut, comme je l'ai fait valoir, faire l'expérience d'une rétroaction qui favorise à la fois l'apprentissage en ce qui concerne son propre bien et le développement de stratégies plus rationnelles. De même, le mécontentement suscité par les déviations quant à la rationalité sociale peut produire une rétroaction qui, à l'échelle sociale, favorise le développement de normes qui se rapprochent davantage de la rationalité sociale. Le potentiel d'agitation sociale qui existe lorsque les intérêts d'un groupe sont mis à l'écart est un potentiel de pression pouvant être exercé par ce groupe – et ses alliés – dans le but d'obtenir une reconnaissance plus complète de leurs intérêts dans la prise de décision sociale et dans les normes socialement inculquées, régissant la prise de décision individuelle. Il s'agit donc d'une pression qui oriente davantage la résolution de conflits dans une direction présumée par la rationalité sociale, puisqu'il s'agit d'une pression qui consiste à donner davantage de poids aux intérêts de celles et ceux qui sont les plus concernés. Une telle pression peut bien évidemment être plus ou moins déterminée ou cohérente. Elle peut se manifester sous la forme d'une expression idéologique bien diversifiée. Elle peut aussi engendrer des conséquences qui sont plus ou moins utiles et favorables à celles et ceux qui exercent cette pression.³³ Des exemples historiques frappants de mobilisation où des groupes exclus ont fait valoir une représentation plus importante de leurs intérêts sont les rébellions contre le système de propriétés féodales, et plus récemment, les mouvements sociaux contre les restrictions à l'égard des pratiques religieuses, du suffrage et d'autres droits civiques, et des négociations collectives.³⁴

Bien évidemment, d'autres mécanismes ont exercé de manière simultanée une influence sur l'évolution des pratiques et des normes sociales, certains desquels ont eu pour conséquence l'effet contraire.³⁵ La question de savoir si les mécanismes qui agissent en faveur de l'inclusion des intérêts exclus prédomineront dépend d'un éventail complexe de facteurs sociaux et historiques. Il serait bête de penser que les normes de n'importe quelle société actuelle à une quelconque étape de l'histoire se rapprochent étroitement de la rationalité sociale, ou bien qu'il y ait une tendance univoque vers un accroissement de rationalité sociale. À l'exemple des mécanismes à l'œuvre dans l'évolution biologique ou dans les économies de marché, les mécanismes décrits ici opèrent dans un « système ouvert » parmi d'autres mécanismes, et rien ne garantit qu'ils conduisent à l'optimalité ou même à un équilibre au sens d'une approche monotone. Les sociétés humaines ne semblent pas avoir été commencées à un état d'équilibre ou à un état qui s'y rapproche, de sorte que la thèse la plus radicale consisterait à affirmer que, à long terme, en l'absence de certains effets exogènes, on pourrait s'attendre à une tendance irrégulièrement séculière vers l'inclusion des intérêts (ou des intérêts représentés par) des groupes sociaux capables d'un certain degré de mobilisation. Mais dans d'autres circonstances, on pourrait s'attendre même à long terme à ce que le contraire se produise. L'esclavage dans le contexte des plantations du Nouveau Monde – assurément, l'une des organisations sociales fondées sur l'exclusion la plus brutale qui n'ait jamais existé – a émergé tardivement dans l'histoire mondiale et a persisté pendant plusieurs siècles. D'autres organisations sociales fondées sur l'exclusion brutale, qu'elles soient anciennes ou récentes, persistent toujours.

Il n'est pas nécessaire, par conséquent, d'épouser une théorie du progrès moral afin de constater que le mécanisme de rétroaction que j'ai décrit jusqu'à maintenant peut accorder un rôle explicatif à la notion de rationalité sociale. L'une des objections les plus déconcertantes, et pourtant la plus courante, adressée au réalisme moral consiste à souligner l'absence de progrès historique uniforme en vue d'un consensus mondial sur les normes morales. Mais l'argument suivant lequel, par exemple, certaines cultures et sous-cultures contemporaines n'acceptent pas – et ne semblent pas être en voie d'accepter – la vision scientifique du monde n'a jamais, à ma connaissance, été invoqué contre le réalisme *scientifique*. Les défenseurs du réalisme scientifique ne sont, bien sûr, pas tenus d'affirmer, dans les deux cas, que des pratiques dans certaines circonstances tendent à produire des théories qui s'appliquent davantage à la réalité; en particulier, lorsque l'objet d'étude est considérablement complexe et si éloigné de la réalité telle qu'elle nous est donnée dans l'inspection directe. Il n'est pas nécessaire qu'ils souscrivent à l'idée désuète voulant que « la vérité éclatera au grand jour » quoiqu'il arrive. La théorie étendue de la rationalité individuelle, par exemple, nous conduit à présager que, dans les sociétés où il existe des conflits d'intérêts considérables, d'une part, les individus développeront des désaccords normatifs importants et, de l'autre, si ces conflits d'intérêts s'établissent en parallèle avec des différences considérables de pouvoir (comme c'est généralement le cas), il est peu probable que les perspectives normatives dominantes soient l'expression de la rationalité sociale. Il est question ici, comme dans le cas de toutes

les explications critérielles en général, d'une explication de certaines tendances [*patterns*] parmi d'autres, et non nécessairement de l'existence d'une seule tendance globale. L'existence des mécanismes de rétroaction, tels que décrits ici, semble toutefois nous offrir des motifs raisonnables de croire que nous pouvons nous rapporter de manière appropriée [*qualified*] à des expériences historiques à titre de quelque chose comme une preuve [*evidence*] expérimentale à propos des types de pratiques dans lesquels un éventail de circonstances pourrait mieux satisfaire à un critère de rationalité sociale. Autrement dit, nous pouvons assigner à ce mécanisme un rôle dans un processus adéquat [*qualified*] d'apprentissage moral.

Les mécanismes d'apprentissage en matière de rationalité individuelle – faible ou étendue – se prêtent à des remarques similaires. Car même si nous nous attendons à ce que, dans des conditions favorables, des individus puissent s'améliorer et mieux agir dans une perspective rationnelle-instrumentale à mesure qu'ils acquièrent plus d'expériences, nous sommes pleinement conscients du fait qu'il existe des mécanismes puissants qui favorisent des résultats opposés. Je ne crois certainement pas qu'un individu doit faire preuve d'une rationalité sans exception, ou même faire preuve d'une rationalité continuellement croissante au cours de sa vie, afin d'appliquer des explications consistant à donner des raisons [*reason-giving explanations*] à plusieurs de ses actions. Je ne pense pas, non plus, que la persistance inévitable de domaines d'irrationalité chez certains individus représente une raison de nier que ceux-ci peuvent, au cours de l'expérience, acquérir des domaines de rationalité plus substantielle.

La comparaison avec la rationalité individuelle ne devrait pas être poussée trop loin. D'abord, bien que les mécanismes qui génèrent l'inclusion dans le cadre de la rationalité sociale agissent par l'intermédiaire de comportements d'individus, des dynamiques interpersonnelles entrent inéluctablement en ligne de compte de telle sorte que les critères sélectionnés ne sont pas réductibles à ceux à l'œuvre dans la rationalité individuelle désagrégée. L'évolution sociale et biologique implique des mécanismes de sélection qui favorisent des comportements satisfaisant aux critères d'optimalité relative, qui sont collectifs (tels que dans le cas du dilemme du prisonnier) ou génotypiques (lesquels peuvent être également collectifs, tel que dans le cadre de la sélection de parentèle), tout aussi bien qu'individuels ou phénotypiques. Si tel n'était pas le cas, il n'aurait guère été possible que des normes morales émergent ou qu'elles aient la prise qu'elles ont désormais sur nous.

Deuxièmement, il existe des différences de degré plutôt extrêmes entre les cas individuels et les cas sociaux. De manière plus frappante, les mécanismes par lesquels les désirs [*wants*] et les comportements individuels sont mis en harmonie avec les intérêts et les raisons individuels agissent de manière plus directe et fiable que les mécanismes analogues qui orientent les pratiques et les normes sociales en direction de ce qui est socialement rationnel. Non seulement les demandes d'information sont moins formidables dans le cas individuel – c'est le moins que l'on puisse affirmer –, mais les façons par lesquelles la rétroaction

est réalisée tendent plus vraisemblablement dans le cas individuel à favoriser le changement et à être moins susceptibles d'être déformées par des asymétries sociales.

Néanmoins, nous avons ici l'esquisse d'une théorie explicative qui emploie la notion de ce qui est plus ou moins rationnel d'un point de vue social et qui, dans le cadre d'explications des croyances et des comportements individuels, reflète très clairement et de manière parallèle le recours à des évaluations de la rationalité du point de vue de l'agent. Tout comme dans la théorie individuelle, elle suggère des généralisations s'appuyant sur des prédictions et des contrefactuels du type suivant : au fil du temps, et dans certaines circonstances plus que dans d'autres, nous devrions nous attendre à ce qu'une pression soit exercée en faveur de pratiques qui satisfont plus adéquatement un critère de rationalité.

Si l'on a ici une théorie potentiellement prédictive et explicative, quelle en est la valeur? Il s'agit là d'une question d'envergure, dont la réponse dépasse le cadre de mes compétences. Mais nous pouvons brièvement noter trois tendances [*patterns*] dans l'évolution des normes morales qui semblent, selon moi, corroborer les prédictions de cette théorie, lesquelles sont soumises aux restrictions liées à l'existence d'imperfections et de mécanismes concurrents. Je les expose toutefois avec une certaine appréhension, car bien que les tendances [*patterns*] que je décris ici soient des tendances historiques grossières, il n'appartient pas nécessairement à la théorie que l'histoire démontre ces tendances, et il ne s'ensuit pas que la théorie doive souscrire à une série de pratiques ou de normes simplement puisqu'elle en résulte.

La généralité. Il s'agit d'un lieu commun en anthropologie que les peuples tribaux ont souvent un terme pour désigner à la fois leur tribu et « le peuple » [*the people*] ou « l'humanité ». Celles et ceux qui ne font pas partie de la tribu ne sont pas considérés comme des individus à part entière, et le type d'obligations que l'on se doit d'avoir auprès des individus ne s'applique pas pleinement aux étrangers. Au cours de l'histoire, par l'entremise de processus qui ont connu de nombreux revirements, les individus se sont accumulés de manière à former des unités sociales plus vastes – de la bande familiale à la tribu, en passant par le « peuple » jusqu'à l'État-nation – et l'étendue des catégories morales s'est élargie afin de suivre cet élargissement des frontières. Il va sans dire qu'il ne s'agit pas là d'une propagation contagieuse des Lumières. Les entités sociales en expansion assujettissent souvent ceux qui se trouvent à l'intérieur de leurs nouvelles frontières, et ceux qui se trouvent dès lors en situation d'oppression parviennent à se garantir une plus grande reconnaissance par des moyens considérablement conflictuels, qui demeurent – et demeureront peut-être toujours – incomplets. Néanmoins, la théorie morale contemporaine et, à un degré qui peut surprendre, le discours moral contemporain en sont venus à rejeter toute forme de limitation au-delà de l'espèce elle-même.³⁶

L'humanisation. Diverses origines et natures ont été assignées aux principes moraux : soit en tant que commandements qui proviennent d'une origine surna-

turelle, fondés dans la volonté ou le caractère d'une déité et devant être interprétés par des prêtres; soit en tant qu'exigences formelles dérivées d'un code d'honneur ancré dans un système de castes; soit en tant que principes d'ordre cosmique; soit en tant que dictats de la raison ou de la conscience, qui ne font appel ni aux inclinations humaines, ni au bien-être, etc. Bien que des vestiges de ces idées survivent dans le cadre de la théorie morale contemporaine, presque toutes les théories morales contemporaines et la majeure partie du discours moral contemporain en général établissent une relation intrinsèque entre les principes normatifs et les effets qu'ils ont sur les intérêts humains. En effet, l'émergence de la moralité en tant que sujet distinct de la religion témoigne de cette tendance.

Les tendances [patterns] de variation. En plus de constater des tendances [patterns] qui reflètent une forme de pression en direction d'une approximation de la rationalité sociale, nous devrions être mesure de remarquer une plus grande approximation de ce genre dans les domaines de régulation normative où les mécanismes postulés ici fonctionnent le mieux, tels que, par exemple, dans les domaines où presque tous les individus ont des intérêts considérablement similaires ou mutuellement dépendants du point de vue de leur satisfaction; où presque tous les individus peuvent éventuellement porter atteinte de manière substantielle aux intérêts d'autrui; où les avantages liés à certaines formes de contrainte ou de coopération sont considérablement saillants même dans le cadre de dynamiques de petits groupes, et où les individus peuvent influencer largement, selon toute vraisemblance, les comportements d'autrui orientés par des normes en obéissant eux-mêmes à des normes. Les exemples les plus évidents ont trait à l'interdiction de l'agression, du vol et des violations de promesses.³⁷ En revanche, il est moins probable qu'une approximation rapide et stable de la rationalité sociale soit réalisée lorsqu'il s'agit de questions morales liées à des situations où il n'existe aucune solution compatible avec la protection des intérêts les plus fondamentaux de tous; où il existe des asymétries très considérables en ce qui concerne la capacité de porter atteinte aux intérêts; où les gains et les pertes qui découlent de certaines formes de coopération ou de contrainte sont difficiles à percevoir; et où la conformité personnelle aura peu d'effet sur la conformité générale. À titre d'exemple, nous pouvons citer des cas qui ont trait à des questions de hiérarchie sociale – par exemple, la permissibilité de l'esclavage, de gouvernements autoritaires, d'inégalités entre les castes et les genres – et de responsabilité sociale – par exemple, quelle est la nature de nos obligations individuelles ou collectives en ce qui concerne la promotion du bien-être des individus avec lesquels nous n'entretenons aucun lien?

Étant donné une caractérisation adéquate des conditions qui ont prévalu au cours des processus d'évolution normative que ces tendances [patterns] décrivent, la présente théorie affirme non seulement que ces changements pouvaient être anticipés, mais qu'une part essentielle de leur explication renvoie à un mécanisme par lequel les individus dont les intérêts ne sont pas pris en considération sont menés à former des valeurs communes et à faire cause commune en vue d'intérêts communs, de sorte qu'ils exercent ainsi une pression sur les pratiques sociales de manière à ce qu'elles en viennent à se rapprocher plus étroitement à la rationalité sociale.

Il ne fait aucun doute que, selon certains, ces descriptions et ces explications de certaines des caractéristiques principales de l'évolution des normes morales semblent, au mieux, naïves, voire, au pire, dangereusement fausses. Cela, je le comprends entièrement. J'ai mis de l'avant des explications douteuses, unilatérales et simplistes à propos d'une réalité très complexe.³⁸ Je ne peux qu'espérer que ces explications semblent aussi crédibles que ce à quoi on peut s'attendre d'explications douteuses, unilatérales et simplistes, et que cela rendrait beaucoup plus plausible l'histoire que j'ai tenté de raconter à propos des mécanismes et des explications.

Il va sans dire que ce qui en résulte n'est pas un fonctionnalisme complaisant ou une approbation globale des pratiques ou des normes morales actuelles. Au contraire, la conception de la moralité esquissée ici met davantage l'accent sur le conflit que sur l'équilibre, et elle nous offre les moyens de critiquer certaines pratiques et intuitions morales contemporaines en posant la question de leur genèse historique. Si nous en venons à penser, par exemple, que l'explication d'une intuition morale commune n'assigne aucun rôle considérable aux mécanismes qui pourraient vraisemblablement exercer une pression en vue de résultats [*outcomes*] socialement rationnels, il s'agit alors d'une raison valable pour remettre en question cette intuition, et ce, malgré le fait que nous y tenons fermement. Dans l'esprit d'une épistémologie morale naturalisée, nous pouvons nous demander si l'explication consistant à élucider la raison pour laquelle nous émettons des jugements moraux constitue un exemple de processus fiable qui nous permet de découvrir des faits moraux.

V. LIMITATIONS

J'ai discuté jusqu'ici de l'excellence morale [*morally best*] sous l'angle de ce qui est rationnel d'un point de vue instrumental et social. Mais j'ai également caractérisé un point de vue de véritablement moral comme ce qui est impartial relativement aux intérêts de tous les individus potentiellement concernés, une notion qui n'est pas socialement conditionnée [*socially-bound*]. En fait, j'ai affirmé qu'une tendance qui va à l'encontre de la spécificité sociale correspond à l'une des tendances [*patterns*] constatables dans le cadre de l'évolution des normes morales. L'explication de ces tendances [*patterns*] – et, par conséquent, le rôle explicatif des degrés de rationalité impartiale – consiste à faire valoir que les mécanismes évoqués ci-dessus ne sont pas également conditionnés sur le plan social [*socially-bound*]. Les sociétés, ainsi que les individus situés aux côtés opposés des frontières sociales, se contraignent mutuellement de diverses manières, tout comme les groupes et les individus se contraignent mutuellement au sein des sociétés : ils peuvent exprimer la menace de commettre une agression, mobiliser une résistance à l'égard des formes de contrôle externe, refuser la coopération et encombrer les plans d'autrui, et ils sont enclins à recourir à ces activités contraignantes lorsque leurs intérêts ne sont pas pris en considération ou sont en danger. Comme pour la moralité intrasociale, les éléments rationnels les mieux établis et presque les plus impartiaux dans le cadre de moralité inter-sociale sont ceux qui sont à l'œuvre lorsque les mécanismes évoqués fonction-

nent de manière optimale [*most reliably*] : les interdictions d'agression sont plus contraignantes et plus largement acceptées que les principes d'équité ou de redistribution. Bien entendu, il y a plusieurs facteurs qui font en sorte que les dynamiques intersociétales diffèrent des dynamiques intrasociétales... Mais j'épargnerai pour une fois le lectorat de toute forme de science sociale de comptoir. Et pourtant, ce qui en ressort est une forme de réalisme moral qui est essentiellement lié à un point de vue limité, à un point vu impartial et pourtant humain. S'agit-il d'une perspective trop limitée pour avoir un réalisme moral authentique?

Un de mes professeurs a déjà souligné que la question du réalisme moral semblait, selon lui, se ramener à la question de savoir si ce que nous faisons importe ou non à l'univers. Puisque nous avons renoncé depuis bien longtemps à croire en l'idée que nous ne sommes pas indifférents au cosmos, il pensait que nous avions par le fait même renoncé au réalisme moral. Je ne peux être d'accord avec ce jugement que si le réalisme moral implique en effet cette idée; on doit alors – avec soulagement plutôt qu'avec tristesse – y renoncer. Toutefois, l'explication proposée ici nous offre un moyen de comprendre comment les valeurs ou les impératifs moraux peuvent être objectifs sans être cosmiques. Ils n'ont pas à être fondés dans rien de plus transcendantal que des faits à propos de l'être humain et de son environnement, des faits concernant les sortes de choses qui nous importent, et en quoi nos façons de vivre ont une incidence sur ces choses.

Or la présente théorie implique une limitation en un tout autre sens, laquelle peut être une source d'inquiétude beaucoup plus importante du point de vue de la théorie morale contemporaine : elle ne donne pas lieu à des impératifs moraux catégoriques au sens où elle nous fournirait une raison d'agir pour tous les agents rationnels indépendamment de leurs désirs contingents. Bien qu'elle soit troublante, cette limitation ne conduit pas au relativisme étant donné que, selon la présente théorie, la motivation rationnelle n'est pas une condition préalable de l'obligation morale. Par exemple, il pourrait être affirmé honnêtement que nous devrions être plus généreux même si une plus grande générosité ne nous aidait pas à promouvoir nos fins existantes ou même à satisfaire nos intérêts objectifs. Il pourrait en être ainsi parce que l'action moralement bien [*right*] que nous devons poser dépend de ce qui est rationnel d'un point de vue qui implique le nôtre, mais sans toutefois se limiter à lui.

De manière similaire, on pourrait affirmer que nous ne devrions pas croire à la fois à une proposition p et à une proposition qui implique non- p . Cependant, il se peut que chaque agent rationnel n'ait aucune raison instrumentale de purger toutes les contradictions logiques de son esprit. Une telle tâche exigerait pour quiconque une part considérable de réflexion si chacun devait tester la cohérence de toutes ses croyances existantes et s'assurer que chacune de ses croyances nouvellement acquises respecte cette cohérence. Supposons que quelqu'un soit si privilégié que les seules contradictions dans ses croyances résident en profondeur dans le marécage considérablement sédimenté de ses

connaissances factuelles d'ordre général. Il se peut que ses souvenirs au sujet de deux connaissances [*acquaintances*] de son passé soient devenus si entremêlés les uns avec les autres que quelque part dans la boue se trouvent deux croyances distinctes qui, mises ensemble, attribuent à un même individu des propriétés logiquement incompatibles. Jusqu'au moment où une telle contradiction surgit de manière explicite en pratique, l'individu ne peut avoir aucune raison d'expliquer ses préoccupations actuelles et d'y remédier après coup, pas plus qu'il n'a de raison de quitter son domicile dans la banlieue du New Jersey pour aller chasser des alligators dans le marais d'Okefenokee, au cas où il pourrait se trouver un jour perdu et sans arme dans les remous de la Géorgie du Sud-Est.³⁹ Ce qu'un individu devrait rationnellement faire peut ainsi différer de ce qui s'impose logiquement à lui. Néanmoins, nous pouvons affirmer que les évaluations logiques ne sont pas subjectives ou arbitraires et qu'il y a de bonnes raisons, parfaitement générales, pour qu'un individu soit logique, notamment que les contradictions logiques sont nécessairement fausses et les inférences logiques sont un gage de vérité. Étant donné que dans le cadre du discours public et dans nos réflexions personnelles nous sommes souvent concernés par la question de savoir si notre pensée est justifiée en un sens qui est lié davantage à ses conditions de vérité qu'à sa valeur instrumentale relative à nos buts personnels, le fait que nous attachions autant d'importance à la logique à titre de standard de critique et de critique de soi est donc loin d'être arbitraire.

Parallèlement, si nous adoptons la conception du bien moral [*moral rightness*] proposée ci-dessus nous pouvons affirmer que les évaluations morales ne sont pas subjectives ou arbitraires, et qu'il existe de bonnes raisons générales pour respecter des obligations [*oughts*] morales, notamment que la conduite morale est rationnelle d'un point de vue impartial. Étant donné que dans le cadre du discours public et dans nos réflexions personnelles nous sommes souvent concernés par la question de savoir si notre conduite est justifiable d'un point de vue général plutôt que d'un point de vue simplement personnel, le fait que nous attachions autant d'importance à la moralité à titre de standard de critique et de critique de soi est donc loin d'être arbitraire.

L'existence de phénomènes tels que la religion et l'idéologie attestent de l'omniprésence et de l'importance de notre souci pour la justification impartiale. Au cours de l'histoire, des individus ont sacrifié leurs intérêts et même leurs vies afin de satisfaire aux exigences des religions et des idéologies qui étaient contraignantes pour eux en partie puisqu'elles prétendaient exprimer un point de vue universel – l'universel – d'un point de vue justificatif. La Rochefoucauld écrit que l'hypocrisie est un hommage que le vice rend à la vertu,⁴⁰ mais « hypocrisie » connote ici le cynisme. Il serait plus approprié d'affirmer que l'idéologie est le respect que les partisans accordent à l'impartialité. La moralité n'est pas alors l'idéologie devenue sincère et générale – une idéologie est intrinsèquement portée à des généralisations sincères [*given to heart-felt generalization*]. La moralité est l'idéologie qui a fait face aux faits.

Nous soupçonnons que l'idée selon laquelle les évaluations morales doivent posséder une force catégorique pour des agents rationnels est en partie attribuable à la crainte que l'autorité de la moralité disparaisse si cette idée était niée. Cela serait effectivement le cas si on tenait à l'idée que les impératifs moraux ne pourraient exister pour quelqu'un s'il n'avait aucune raison d'y obéir, car un individu pourrait alors échapper de manière expéditive aux devoirs moraux par le fait d'avoir des désirs bas. Mais si nous abandonnons cette idée concernant l'applicabilité des jugements moraux, des variations dans les désirs personnels ne peuvent alors cautionner aucune exemption sur le plan des obligations morales.⁴¹

Par conséquent, bien que l'argument présenté ici ne parvienne pas à formuler une conception des impératifs moraux en tant qu'impératifs catégoriques, il s'agit là d'une limitation avec laquelle nous pouvons composer, et qui peut néanmoins consacrer à la moralité l'étendue et la dignité dont elle jouit traditionnellement. Car combien parmi nous peuvent parvenir à se convaincre que la raison ne peut être autre qu'hypothétique? Doit-on également poser la question de savoir combien d'individus parmi nous considéreraient que l'importance que nous accordons à la moralité ou à la conduite morale serait augmentée par une démonstration suivant laquelle même une personne ayant les fins les plus répugnantes trouvera que la conduite morale parvient à les promouvoir?

Une des implications de la théorie présentée ici est que si nous voulons que la moralité soit prise au sérieux et qu'elle ait une place importante dans la vie des individus – et non pas simplement en tant qu'illusion ou sous l'effet d'une menace de répression – nous devrions alors nous soucier considérablement des diverses modalités par lesquelles les arrangements sociaux produisent des conflits d'intérêts et des asymétries de pouvoir, qui ont une incidence sur la nature et la taille de l'écart qui sépare la rationalité individuelle et la rationalité sociale. Au lieu de tenter de concevoir la moralité comme ce qu'elle ne peut être, c'est-à-dire comme « ce qui est rationnellement contraignant peu importe les fins de tout et chacun », nous devrions plutôt nous demander comment nous pouvons changer nos modes de vie de manière à ce que la conduite morale soit plus régulièrement rationnelle étant donné les fins que nous avons actuellement.

VI. RÉSUMÉ ET CONCLUSION

J'ai esquissé une forme de réalisme moral et indiqué comment elle peut être défendue contre certaines objections. Je ne peux présenter dans le cadre du présent essai ni une caractérisation complète de cette théorie, ni des réponses complètes aux nombreuses objections qu'elle suscite. Je devrais peut-être alors cesser de me prononcer sur ces questions et clore cet essai en indiquant approximativement ce que j'ai tenté de présenter et ce que je n'ai pas exposé.

J'ai proposé de réformer les définitions naturalistes de ce qui est non moralement bon [*goodness*] et de ce qui est moralement bien [*moral rightness*]. On pourrait me répondre : « Oui, nous pouvons constater qu'une telle ou telle fin est un inté-

rêt objectif de l'agent au sens où vous l'entendez, ou qu'une telle ou telle pratique est rationnelle d'un point de vue impartial, mais pouvons-nous néanmoins nous demander si la fin est bien pour elle ou lui ou si la pratique est bonne [*right*]? » Or de telles « questions ouvertes » ne peuvent de par leur nature être résolues, étant donné que les définitions ne font pas l'objet de démonstrations ou de réfutations. Mais de telles questions ouvertes peuvent être plus ou moins troublantes, car bien que des propositions définitoires ne puissent être démontrées, elles peuvent mieux s'en tirer ou non à la lumière de divers *desiderata*.

J'ai présupposé que le travail d'élaboration de définitions consiste en une partie de la construction théorique et se doit d'être évalué en se demandant 1) si les analyses satisfont aux contraintes adéquates d'intelligibilité et de fonctionnalité, et 2) si les termes en tant qu'ils sont analysés contribuent à la formulation et à la vérification des théories recevables. En quoi notre théorie répond-elle à ces critères?

1) Outre les contraintes d'intelligibilité, telles que la clarté et la non-circularité, les définitions spécifiquement naturalistes de termes évaluatifs doivent satisfaire deux autres contraintes qui relèvent de leurs fonctions spécifiques. a) Elles doivent dans la mesure du possible saisir la force normative de ces termes en proposant des analyses qui permettent à ces termes de jouer leurs rôles évaluatifs fondamentaux. Dans le cadre actuel, cela signifie de démontrer qu'elles expriment néanmoins des notions de ce qui est bon [*goodness*] et de ce qui est bien [*rightness*] que l'on peut reconnaître, bien que les définitions proposées ne correspondent pas à toutes nos intuitions linguistiques et morales. Par ailleurs, cela signifie de démontrer que les définitions permettent d'établir des liens plausibles entre, d'une part, ce qui est bon ou bien et, de l'autre, ce qui typiquement motiverait les individus prêts à se soumettre à des formes adéquates d'examen approfondi. b) Les définitions naturalistes doivent s'assurer que les concepts évaluatifs participent, de plein droit, à des théories véritablement empiriques. Il s'agit, d'une part, de montrer que nous avons un accès adéquatement épistémique à ces concepts. Il s'agit, d'autre part (et corrélativement), de montrer que les généralisations recourant à ces concepts, entre autres, peuvent intervenir dans le cadre de théories potentiellement explicatives. J'ai tenté de proposer des définitions raisonnablement claires et de montrer à titre préliminaire comment elles peuvent satisfaire les contraintes a) et b).

2) Il reste toutefois encore beaucoup à faire, car il faut encore démontrer que les théories empiriques élaborées à l'aide de ces définitions sont des théories raisonnablement bonnes, c'est-à-dire des théories qui s'appuient sur des preuves substantielles et qui donnent lieu à des explications plausibles. C'est ce que j'ai tenté de suggérer à titre d'étape strictement préliminaire. Si j'ai été tout à fait peu convaincant sur les sujets d'ordre empirique, je m'attends alors à ce que les définitions proposées ici soient aussi peu convaincantes.

Un atout du naturalisme en éthique et en épistémologie tient, selon moi, au fait qu'il présente des contraintes dans plusieurs domaines importants à la fois.

Plusieurs occasions s'offrent à nous pour démontrer que la théorie est fautive ou est jugée peu convaincante lorsqu'elle doit répondre à des exigences empiriques ainsi qu'à des intuitions normatives. L'effort de théorisation en général s'avère plus productif lorsque des contraintes adéquates s'imposent; en particulier, des contraintes sont exigées en éthique si nous devons avoir une idée plus claire quant à la façon de faire des progrès en ce qui concerne la résolution théorique de conflits. Bien sûr, il ne s'agit pas simplement de n'importe quelle contrainte. L'ensemble des contraintes proposées doit se présenter à la fois comme adéquat et utile. Pour conclure, nous ferons quelques remarques au sujet de 1) l'utilité des contraintes adoptées ici et 2) sur leur pertinence.

1) Considérons trois classes de théories qui concurrencent la théorie morale substantielle qui a été exposée ci-dessus, et remarquons comment les critiques qu'on leur adresse confondent naturellement des préoccupations relatives à la justification normative et à l'explication empirique. On considère généralement que les conceptions *kantiennes* de la moralité rendent compte de certaines caractéristiques normatives, intuitivement probantes au sujet de notions telles que la rationalité et le bien moral [*moral rightness*], mais il semble que c'est possible, en partie, par l'omission de l'élaboration d'un moyen plausible par lequel ces notions seraient intégrées dans une théorie empirique des raisons et des motifs d'agir. Qui plus est, cette difficulté descriptive trouve son expression directe dans le domaine normatif. Non seulement une « obligation » [*ought*] normative doit relever d'une « capacité » [*can*] empirique, mais une obligation normativement convaincante [*compelling*] doit – comme l'ont souligné les critiques récentes du kantisme – atteindre les véritables sources de l'action et des préoccupations humaines. Les théories morales *intuitionnistes* ont également joui d'un certain succès par rapport au fait de saisir certaines caractéristiques normatives de la moralité, mais elles ont considérablement été abandonnées faute de fournir une théorie crédible de la nature et de l'opération de la faculté de l'intuition morale. Il nous est trop facile de formuler une explication psychologique non justificative de l'existence en de quelconques gentlemen anglais de ce que ceux-ci parviennent à identifier au moyen de l'introspection comme une faculté de l'intuition morale, une explication qui lie cette prétendue faculté plus étroitement avec la rigidité des conventions sociales existantes qu'avec quelque chose qui pourrait correspondre en apparence à une source de vérité universelle.

2) L'enjeu qui se pose toutefois lorsque nous évaluons des approches concurrentes de la moralité implique non seulement une forme d'évaluation des théories importantes, mais également des questions concernant le type de critères d'évaluation qui s'appliquent aux définitions et aux théories en éthique, de même que le fait de déterminer si l'effort considérable de systématisation définitionnelle et de théorisation est tout simplement approprié en ce qui concerne l'éthique. Je suis porté vers l'idée selon laquelle le développement des théories en éthique n'est pas une construction artificielle des philosophes, mais le résultat organique d'usages personnels et sociaux de l'évaluation morale : les individus et les groupes ont fait face, maintes et maintes fois, à des questions difficiles auxquelles le sens commun ne pouvait que donner des réponses

conflictuelles ou sinon insatisfaisantes, de sorte qu'ils ont approfondi davantage leurs questions et ont poursuivi leur enquête de manière plus systématique. Ils ont ressenti le besoin de formuler une théorie en éthique qui s'articulerait en parallèle avec le besoin éprouvé d'élaborer une théorie relevant des sciences naturelles ou sociales.⁴² Il n'en résulte pas que seule la théorisation éthique doit s'articuler en parallèle avec – ou s'intégrer dans – la théorisation dans le cadre des sciences naturelles et sociales. Bien qu'elle puisse sembler initialement plausible et ultimement irréfutable, l'idée suivant laquelle l'éthique doit être isolée est également une idée que nous rejetons ultimement. Nous sommes des êtres naturels et sociaux, et nous n'avons aucune idée où nous pourrions regarder ailleurs, en ce qui concerne l'éthique, que dans cette riche conjonction de faits. J'ai tenté de suggérer que c'est dans ces domaines que nous pourrions effectivement trouver une éthique.⁴³

REMERCIEMENTS

Traduction de Peter Railton, “Moral Realism”, *The Philosophical Review*, Vol. 95, No. 2, p.163-207, par Denis Courville. *Les ateliers de l'éthique/The Ethics Forum* souhaite remercier *Duke University Press* de nous avoir accordé la permission de traduire et de publier ce texte ainsi que Denis Courville pour en avoir fait la traduction.

NOTES

- ¹ Le naturalisme évolutionnaire du XIX^e siècle offre un exemple de la première variante, alors que Dewey – et, du moins selon une certaine interprétation, peut-être Mill également – offre un exemple de la seconde.
- ² A. J. Ayer, « Langage, vérité et logique », Paris, Flammarion, 1956, p. 156-157.
- ³ En faisant valoir cette idée, je soutiens que les questions portant sur ce qui est rationnel de croire appartiennent à la raison pratique plutôt qu'à la raison théorique. Bien que les résultats de la raison théorique – par exemple, les conclusions d'inférences déductives – soient en général pertinents pour les questions à propos des croyances rationnelles, ils ne sont pas déterminants, mis à part les raisons pratiques d'un agent.
- ⁴ Bien entendu, la formation de croyances individuelle n'est pas en général gouvernée explicitement par un raisonnement de type moyen-fin, mais plutôt par des habitudes liées à la formation de croyances et par des tendances à investir des degrés variables de confiance dans certaines croyances. Si nous acceptons une conception instrumentale de la rationalité, nous pouvons alors qualifier de rationnelles ces habitudes du point de vue de l'individu, dans la mesure où elles s'insèrent dans une constellation d'attitudes et de tendances qui font la promotion de ses fins. Cette question sera abordée de nouveau dans la section IV.
- ⁵ Ni ces remarques, ni celles formulées dans les paragraphes suivants se veulent une exégèse sérieuse des arguments de Hume, lesquels admettent des interprétations autres que ceux qui sont offerts ici. Je désire seulement rendre compte de certaines caractéristiques de ce que je considère être les arguments de Hume, par exemple, dans le livre III, partie I, section I du *Traité de la nature humaine (La morale)*, Paris, Flammarion, 1993, en particulier p. 60-61, et dans l'annexe I de *l'Enquête sur les principes de la morale*, Paris, Flammarion, en particulier p. 214-215.
- ⁶ Philippa Foot a remis en cause cette thèse, et bien que la façon dont elle pose et fait valoir la question diffère suffisamment de la mienne que je ne puis juger si elle est d'accord avec l'argument qui suit. Voir son ouvrage *Virtues and Vices*, Berkeley, University of California Press, 1978; en particulier l'essai XI. La présentation de cette question s'inspire principalement de la distinction établie par William K. Frankena entre le point de vue rationnel et le point de vue moral.
- ⁷ Voir Hume, David, *Enquête sur les principes de la morale*, section IX, partie II, p. 196.
- ⁸ La critique ancienne du non-cognitivism, selon laquelle il éprouve de la difficulté à rendre compte de la différence entre les valeurs morales et les autres types de désirabilité (de sorte que Hume peut parler dans un même souffle de « bons offices » et de son logis bien disposé) connaît un certain regain de vitalité dans le présent contexte. Il est nécessaire pour rendre compte de différences de ce genre que l'on dispose d'un moyen substantiel [*contentful*] de caractériser des critères d'évaluation morale afin que l'approbation morale ne soit pas réduite à « ce à quoi un agent accorde de l'importance ». (Une caractérisation de ce type sera présentée à la section IV.) Une valeur *sans phrase* [en français dans le texte] est une notion générique, et non nécessairement morale. Il est parfois affirmé que la valeur générique présente un caractère moral lorsque l'on atteint ce à quoi un agent accorde par-dessus tout de l'importance. Mais cela aurait pour effet d'investir de valeur morale les animaux de compagnie et les mementos, et aurait l'effet particulier de rendre presque impossible l'amoralisme. Il semble plus plau-

sible d'affirmer que les valeurs ne sont pas toutes morales, et que les valeurs les plus élevées ne doivent pas nécessairement être – ou même simplement être considérées par un agent comme – des valeurs morales. Lorsque l'on se tourne vers des questions d'obligation [*duty*], la situation devrait être encore plus claire : les théoriciens moraux ont proposé des relations tout à fait différentes entre les catégories à propos de ce qui est moralement bien [*moral rightness*], moralement bon [*moral goodness*], et non moralement bon [*non-moral goodness*], et il ne semble pas plausible d'affirmer que le fait de considérer une action ou une classe d'actions comme moralement bonne [*right*] est nécessairement équivalent au fait de la considérer personnellement comme ce qui a de valeur [*valuable*] sans phrase.

⁹ La nature des impératifs moraux est discutée plus longuement à la section V.

¹⁰ Des théories de ce type sont dûment externalistes lorsque, en définissant les notions de *fiabilité* [*reliability*] ou de *processus causaux qui confèrent la justification* [*warrant-conferring causal process*], elles ont recours à une théorie de la vérité qui ne réduit pas celle-ci à ce en vertu de quoi nous avons des raisons de croire – par exemple, une théorie correspondantiste non triviale.

¹¹ Ou si les faits moraux étaient censés d'être le type de choses qui fournit des raisons catégoriques d'agir. Toutefois, cette supposition consiste simplement à revêtir la thèse humienne de la praticabilité d'une tenue ontologique.

¹² J. L. Mackie, in *Ethics: Inventing Right and Wrong*, Harmondsworth, Middlesex, Penguin, 1977; et Gilbert Harman, in *The Nature of Morality: An Introduction to Ethics*, New York, Oxford University Press, 1977. Tous les deux contestent le réalisme moral en partie en remettant en cause sa capacité d'explication. Nicholas L. Sturgeon, in « Moral Explanations », David Copp and David Zimmerman (eds.), *Morality, Reason and Truth: New Essays in the Foundations of Ethics*, Totowa, N.J.: Rowman and Allanhead, 1984, adoptent l'opinion contraire, en ayant recours à des arguments distincts de ceux présentés ci-dessous.

¹³ Je crois qu'une théorie exhaustive de la valeur pourrait démontrer que le concept du bien [*good*] de quelqu'un est légèrement différent du concept de ce qui est désirable pour lui. Cette différence n'aura toutefois aucune incidence sur l'argument présenté ici.

¹⁴ C'était le travail de Richard C. Jeffrey sur les probabilités épistémiques qui m'ont conduit à la base à l'idée d'intérêts subjectifs objectifiés. Voir la note 17. J'ai tiré profit depuis lors du travail de Richard B. Brandt sur les « désirs rationnels », bien que je craigne qu'il considère une part notable de choses affirmées ici comme erronée. Voir la partie I de *A Theory of the Good and the Right*, Oxford, Clarendon, 1979.

¹⁵ Nous posons cette question à *A+*, plutôt que celle de savoir ce que *A+* voudrait pour lui-même, puisque nous sommes à la recherche des intérêts subjectifs objectifiés de *A*, et les intérêts de *A+* pourrait être tout à fait différents en raison des changements impliqués dans l'idéalisation de *A*. Par exemple, on peut supposer vraisemblablement que *A+* ne désire pas davantage d'informations au sujet de lui-même, car il n'y a rien de plus à ajouter, et cela il le sait. Il pourrait néanmoins être tout à fait possible que *A+* désirera davantage d'informations s'il était à la place de la version moins informée de lui-même, à savoir *A*. Cela pourrait être impossible, en tant que question psychologique, pour *A+* de mettre complètement de côté ses désirs dans ses circonstances actuelles au sujet de lui-même ou *A*, lorsqu'il s'agit de considérer ce qu'il voudrait vouloir s'il se trouvait à la place de la version moins qu'idéale de lui-même. Tout cela révèle un problème de mesure concernant les intérêts objectifs : le fait de mettre à la disposition d'un individu les informations et les capacités nécessaires à l'« objectivation » de ses intérêts peut perturber sa psychologie de manière à ce que le phénomène que l'on désire observer s'en trouve modifié. De telles difficultés se présentent même au cours de la mesure des intérêts subjectifs étant donné que les instruments utilisés pour échantillonner les préférences (en effet, de simples actes de réflexion d'un individu sur ses propres préférences) tendent à avoir une incidence sur les préférences exprimées. Pour des raisons évidentes, les effets d'interférence sont inévitables. Bien qu'ils ne constituent pas en eux-mêmes une raison suffisante pour justifier une forme de scepticisme à l'égard des intérêts subjectifs et objectifs, ces problèmes de mesure illustrent bien l'importance d'une « théorie des perturbations » [*perturbation theory*], et la nécessité de faire preuve de prudence à l'égard des attributions d'intérêts qui se révèlent inattentives aux effets d'interférence.

- ¹⁶ « Intérêt » n'est pas tout à fait le mot recherché ici, car dans le langage courant nous pouvons parler d'un désir (*want*) là où nous ne pourrions pas parler d'un intérêt correspondant. Cf. Brian Barry, *Political Argument*, Londres, Routledge et Kegan Paul, 1965, en particulier, le chapitre X, pour une discussion de ce genre. Une expression plus précise, mais excessivement encombrante, serait « caractéristique donnant lieu à une valeur positive » [*positive-valence-making characteristic*].
- ¹⁷ Supposons un instant, contrairement à ce qui a été établi ci-dessus, que nous disposons d'une notion acceptable de probabilité épistémique qui détermine des degrés rationnels de croyance indépendamment des objectifs contingents de l'agent épistémique. Peut-être que l'analogie suivante sera alors utile. Prenons un processus physiquement contingent, comme la désintégration alpha. Nous pouvons demander à un individu quelle probabilité subjective il assignerait à un événement consistant dans une certaine vitesse [*rate*] de dégradation pour un quelconque échantillon d'uranium; nous pourrions aussi demander quel est le degré rationnel de croyance qu'un individu pourrait assigner à cet événement s'il devait alors être idéalement informé des lois de la physique et des conditions initiales pertinentes. Nommons ce degré rationnel de croyance la probabilité subjective objectivée de l'événement, et supposons qu'elle soit égale à un cinquième (à comparer avec : Jeffrey, Richard C., *The Logic of Decision*, New York, McGraw-Hill, 1964, p. 190-196). Mais considérons à présent les faits physiques qui, en lien avec les lois de la mécanique quantique, fondent le jugement idéalisé de l'individu. Appelons-les la base réductive de ce jugement. Cette base réductive consiste dans un ensemble complexe de qualités premières, dont on peut dire qu'elle rend possible le fait qu'à l'événement en question correspond une probabilité objective d'un cinquième (il faut préciser que l'approche de Jeffrey ne postule pas de telles probabilités objectives.) L'existence de cette probabilité objective peut expliquer la raison pour laquelle un individu idéalement informé pourrait sélectionner une probabilité égale à un cinquième, mais le jugement de probabilité d'un individu idéalement informé ne peut expliquer pourquoi la probabilité objective est d'un cinquième, car cela relève des lois de la physique. De même, l'existence d'un intérêt objectif d'un individu peut expliquer la raison pour laquelle la version idéalement informée de cet individu pourrait sélectionner un intérêt subjectif objectivé pour la version moins informée de lui-même, mais non l'inverse.
- ¹⁸ Plus précisément, nous pourrions affirmer que X est non moralement bon [*good*] pour A à un temps t si et si seulement X satisfait un intérêt objectif de A lorsqu'une base réductive existe à t . Un bon nombre de questions peuvent être soulevées au sujet des considérations relatives à l'évolution des intérêts dans le temps, mais on ne peut les aborder ici.
- ¹⁹ A^+ , en se mettant à la place de A , peut trouver plusieurs ensembles de désirs [*wants*] différents tout aussi attrayants les uns que les autres, de telle sorte que plusieurs options peuvent être, en ce sens, équivalentes-meilleures [*equal-best*] pour A . Sans pour autant devenir problématique, la notion du « bien pour A » deviendrait alors plutôt pluraliste [*pluralistic*]. Toutefois, une question plus sérieuse se pose : la spécification des conditions de A^+ ou de la psychologie du désir sont-elles assez déterminées pour que la notion d'intérêt objectif soit suffisamment définie afin de répondre à mes objectifs? Sans préciser quel pourrait être ce niveau de caractérisation, je propose deux façons d'esquisser une réponse au problème de la précision [*definiteness*]. 1) Il semble que nous pensons effectivement qu'il existe des réponses plus précises quant à la façon dont les désirs d'un individu A pourraient être modifiés si ses croyances devaient également se modifier sous certaines conditions. Si Lonnie devait apprendre ce qui résulte du fait de boire du lait, il ne voudrait plus que son désir de boire du lait soit opérant. Mais un changement considérable dans les croyances peut s'effectuer petit à petit sur la base d'une séquence de changements limités dans les croyances. Ainsi, si (et il s'agit là, certes, d'un grand « si ») l'ordre du changement n'a aucune importance au final, alors les faits et les généralisations sur lesquels s'appuient les contrefactuels à propos des changements limités pourraient appuyer une extrapolation jusqu'à A^+ . 2) Les croyances et les désirs semblent être sujets à une covariation systématique. En général, nous constatons que des individus qui diffèrent notablement dans leurs désirs – par exemple, concernant leurs carrières ou leur style

de vie – diffèrent notablement, et typiquement, dans leurs croyances; au fur et à mesure que les individus se ressemblent dans leurs croyances, ils tendent à se ressembler dans leurs désirs. Cela laisse à penser que si (un autre grand « si ») la caractérisation proposée de A^+ fixe de part en part le contenu de ses croyances de manière définitive (du moins, étant donné un choix de langage), alors ses désirs peuvent également être complètement fixés. Si nous disposions d'une théorie générale de la covariation des croyances et des désirs, nous pourrions alors faire appel directement à cette théorie – en plus des faits à propos de A – pour déterminer les contrefactuels nécessaires pour caractériser les intérêts objectifiés de A , de sorte que nous pourrions éliminer par là même toute référence essentielle à l'individu imaginaire A^+ .

²⁰L'explication en question peut, toutefois, donner lieu à des résultats contre-intuitifs. Suivant la nature et les circonstances d'individus donnés, les intérêts objectifs de ceux-ci peuvent consister dans des choses que nous considérons comme mal et répugnantes, et qui ne semblent pas, selon nous, faire partie de ce qu'est une bonne vie. Nous pouvons expliquer une bonne part de nos objections à certains désirs – par exemple, ceux qui impliquent la cruauté – en affirmant qu'ils ne sont pas *moralement* bons [*good*]; d'autres désirs – par exemple, de nature philistine – en affirmant que, d'un point de vue *esthétique*, ils n'ont aucune valeur, et ainsi de suite. Il me semble préférable d'exprimer notre dégoût pour certaines fins en termes de catégories spécifiques de valeur plutôt que d'avoir recours à l'idée suivant laquelle de telles fins ne pourraient en aucune circonstance faire partie du bien non moral [*non-moral good*] de quelqu'un. L'ensemble des personnes – ou du moins certains individus – pourrait être organisé de manière à ce que certaines choses qui ne semblent pas très attrayantes deviennent essentielles à son épanouissement [*flourishing*], et nous ne voulons pas être coupables de vœux pieux à ce sujet (il y aura suffisamment de vœux pieux d'ici la fin).

²¹Pour reformuler la question en des termes plus strictement naturalistes, nous pourrions nous attendre à ce que l'évolution ait favorisé des organismes constitués de telle sorte que les comportements nécessaires à leur survie et à leur développement [*flourishing*] soient associés à des états internes positifs (tels que le plaisir), alors que les comportements qui s'opposent à leur survie et à leur développement [*flourishing*] soient associés à des états négatifs (tels que la douleur). La notion de « développement » [*flourishing*], même si elle est comprise comme une simple aptitude à la survie [*fitness*] reproductive, n'est pas une notion étroite. Pour que des êtres tels que les êtres humains puissent être reproductivement féconds [*successful*], ils doivent, en tant que phénotypes, avoir des vies qui sont psychologiquement viables [*sustainable*], motivées de manière interne [*internally motivating*], et effectivement sociales. Il s'agirait également de vies qui, en général, se livreraient à un grand nombre de capacités spécifiquement humaines. L'espèce humaine ne pourrait difficilement avoir connu le succès qu'elle a connu, même sur le plan reproductif, si ce n'était pas de la recherche [*pursuit*] des types de choses qui, de manière caractéristique, ont incité les êtres humains à agir en fonction de ce type d'existence. Toutefois, il faut garder à l'esprit que la majeure partie du processus de l'évolution humaine a eu lieu dans des circonstances qui se distinguaient, en de multiples façons, de celles d'aujourd'hui. Il est donc tout à fait possible que l'interaction entre le potentiel motivationnel de l'humanité ait évolué et que les circonstances actuelles donneront lieu, à leur tour, à des incongruités entre ce vers quoi nous tendons, ou ce qui nous anime, et ce qui pourrait produire le plus grand plaisir pour nous tous. Il s'agit là d'une raison pour remettre en doute l'hédonisme en tant que théorie de la motivation.

²²Il est question ici de « succès fonctionnel » plutôt que de « précision représentationnelle » [*representational accuracy*] pour les raisons suivantes. La sélection favorise les organismes qui ont telle ou telle autre caractéristique, qui contribue à répondre à leurs besoins dans le cadre d'environnements particuliers. La question de savoir si à cette caractéristique correspond une capacité représentationnelle précise [*accurate*] ne peut être résolue par un argument de ce type. Ce serait, bien sûr, une très grande coïncidence si des êtres qui se fient aussi considérablement que nous sur des représentations étaient seulement aptes à construire des représentations grossièrement inexactes, tout en étant capables de composer fructueusement [*successfully*] avec divers environnements sur l'étendue d'une longue période de temps. Mais de telles coïncidences ne peuvent être écartées.

- ²³ Bien que certains éléments de leur base réductive dépendent de nos choix antérieurs, nos intérêts objectifs ne sont pas, par conséquent, subjectifs en un sens qui est préjudiciable à l'argument présenté ici. Après tout, des faits objectifs qui nous concernent et qui sont aussi non problématiques que notre poids, notre revenu et notre position spatiale dépendent de nos choix antérieurs. Il ne s'agit pas ici d'affirmer que nos intérêts subjectifs n'ont aucun rôle dans la constitution de la base réductive de nos intérêts objectifs, mais plutôt que ceux-ci ne peuvent influencer nos intérêts objectifs qu'en vertu des effets actuels (plutôt que des effets simplement désirés) qu'ils ont sur cette base réductive, tout comme nos choix antérieurs ne peuvent influencer notre poids, notre revenu ou notre position spatiale qu'en vertu des effets actuels (plutôt que des effets simplement désirés) qu'ils ont sur notre déplacement, notre situation d'emploi ou nos mouvements.
- ²⁴ De même, une explication serait incomplète et superficielle si elle tentait d'élucider les raisons pour lesquelles la consommation mondiale de sucre raffiné a connu une croissance fulgurante lorsque la production à grande échelle est devenue possible, en se limitant tout simplement au constat que les gens apprécient le goût du sucre. Pourquoi, malgré la présence d'une forte divergence dans les alimentations traditionnelles et le développement des goûts, le sucre est-il parvenu à faire de si grandes percées dans la consommation humaine? Pourquoi l'apparition et la promotion de produits alimentaires tout aussi peu coûteux n'ont-elles pas provoqué des changements de consommation aussi remarquables? Pourquoi la consommation de sucres, souvent dissimulés sous d'autres formes, continue à croître, et ce, même dans les sociétés où le sucre est reconnu comme présentant un risque pour la santé? Les faits concernant la manière dont nous sommes constitués, les divers moyens singuliers par lesquels le sucre nous affecte, et les façons dont les formes de production et les types de consommation ont conjointement évolué ensemble de manière à générer une demande en croissance et une offre en expansion, doivent compléter une explication qui s'en tient seulement au niveau des préférences subjectives. Pour une discussion approfondie de cette question, cf. Sidney W. Mintz, *Sweetness and Power: The Place of Sugar in Modern History*, New York, Viking, 1985.
- ²⁵ Pour une explication de la façon dont les théories informellement exprimées peuvent néanmoins transmettre de l'information explicative, cf. la section II de notre article « Probability, Explanation, and Information », *Synthese*, vol. 48, 1981, p. 233-256.
- ²⁶ Une explication de ce type emploie un critère naturalisé lorsque la rationalité est définie en termes d'efficacité relative compte tenu des croyances et des désirs de l'agent. Un agent (plus ou moins) rationnel est donc quelqu'un qui est disposé à agir de manière (plus ou moins) efficace. Le fait de qualifier de « naturaliste » une telle explication comprend toutefois son lot de difficultés considérables, car les contraintes auxquelles sont soumises les attributions de croyances et de désirs au moyen d'un « principe de charité » peuvent compromettre l'idée suivant laquelle les explications qui recourent à la notion d'agent rationnel sont empiriques. Bien que je croie que cette difficulté puisse être surmontée, l'argument ne peut être présenté ici.
- ²⁷ Pour une vue d'ensemble de la littérature, cf. Richard Nisbett and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs, Prentice-Hall, 1980. Il n'est pas surprenant que l'on y trouve une attention plus particulière accordée aux inconvénients qu'aux avantages.
- ²⁸ Pour rappeler un point de la section II : on peut procéder à des évaluations relativement à des critères évaluatifs particuliers sans pour autant établir ce qui les satisfait.
- ²⁹ Je suis conscient du fait que la caractérisation d'« un point de vue qui est à la fois impartial et prend en considération tous les individus qui pourraient être éventuellement concernés par cette évaluation morale » en tant que point de vue *social* peut induire en erreur – certains individus éventuellement concernés peuvent se trouver de l'autre côté d'une frontière intersociale. Cette complication sera mise de côté jusqu'à la section V.
- ³⁰ Cela signifie également que la relation entre les critères moraux et les critères de rationalité individuelle devient alors problématique puisque rien ne pourrait garantir que ce qui, du point de vue d'un quelconque individu, serait instrumentalement rationnel coïnciderait avec ce qui, d'un point de vue social, serait instrumentalement rationnel.

³¹ Une thèse plutôt forte à propos des comparaisons interpersonnelles est requise ici à des fins d'agrégation sociale. Je ne présume pas l'existence d'un seul bien, comme le bonheur [*happiness*], qui sous-tendrait de telles comparaisons. Ainsi, la théorie morale en question, bien que conséquentialiste, agrégative et maximisante, ne peut se réduire à l'utilitarisme classique. Je présume que lorsqu'il est question de faire un choix entre satisfaire un intérêt *X* de *A* par opposition au fait de satisfaire l'intérêt *Y* de *B*, les réponses à la question : « m'importerait-il davantage, toutes choses égales par ailleurs, d'avoir *X* satisfait si j'étais *A* plutôt que d'avoir *Y* satisfait si j'étais *B*? », seront relativement déterminées et stables parmi les individus dans des conditions où l'on dispose d'informations vives et complètes. Une forme similaire, mais un peu plus faible, de compatibilité parmi les différences est présumée lorsque nous faisons des choix parmi des lignes d'action alternatives, qui nous conduiraient à avoir des désirs différents à l'avenir.

³² Voir par exemple, Barrington Moore, Jr., *The Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*, Boston, Beacon, 1966, et *Injustice: The Social Bases of Obedience and Revolt*, White Plains, N.Y., M. E. Sharpe, 1978; E. P. Thompson, *The Making of the English Working Class*, New York, Pantheon, 1963; William B. Taylor, *Drinking, Homicide, and Rebellion in Colonial Mexican Villages*, Stanford, Stanford University Press, 1979; Charles Tilly, *From Mobilization to Revolution*, Reading, Mass., Addison-Wesley, 1978; et Charles Tilly et al., *The Rebellious Century, 1830-1930*, Cambridge, Harvard University Press, 1975.

³³ Un thème commun que l'on retrouve dans les travaux cités à la note 32 consiste dans l'idée suivant laquelle une grande part d'agitation sociale est revendicative plutôt que révolutionnaire, étant donné que le mécontentement des groupes assaillis depuis longtemps mène très souvent à l'action et à la mobilisation seulement lorsque les droits coutumiers sont menacés ou niés. Ainsi, les idéologies manifestes de tels groupes sont fréquemment particularistes et conservatrices, tant même que leur agitation contribue à l'émergence de nouvelles formes sociales qui accordent un plus grand poids aux intérêts antérieurement mis à l'écart. De même, des individus qui ne parviennent pas à remarquer l'irrationalité de leur comportement coutumier jusqu'à ce qu'ils y soient conduits par la voie de difficultés non coutumières, ce qui éveille ensuite le sentiment que quelque chose ne va pas. Pour des raisons bien connues, une réponse individuelle initiale consiste généralement à renouer avec le *statu quo ante*, bien qu'il soit possible que des changements authentiques puissent avoir lieu à la suite de ces efforts de restauration.

³⁴ On doit souligner que ces mécanismes ne présupposent pas l'existence préalable d'institutions démocratiques. Ils ont pu extraire des concessions même au sein de sociétés qui sont demeurées profondément hiérarchiques. Voir, par exemple, Taylor, *Drinking, Homicide, and Rebellion*.

³⁵ En effet, le mécanisme décrit ici peut conduire simultanément à divers résultats : vers l'inclusion de certains intérêts antérieurement mis à l'écart, et vers l'exclusion de certains intérêts dont on tenait compte antérieurement. Pour en être certain, si des intérêts en viennent à être mis à l'écart même si leur base matérielle et sociale demeure plus ou moins intacte, un nouveau potentiel d'agitation sociale est créé. Certains groupes représentent alors un problème particulier en raison de leur incapacité inhérente à se mobiliser effectivement, tels que, par exemple, les enfants et les générations futures. Pour rendre compte des pressions qui ont été exercées au nom de ces groupes, il est nécessaire de rendre compte de la façon dont les individus en viennent à comprendre d'autres individus en fonction de leurs propres intérêts. (Comparons les façons par lesquelles nos diverses conceptions de nous-mêmes dans l'avenir, qui ne peuvent exercer aucune pression en elles-mêmes, en viennent à être prises en considération dans le moment présent en raison du fait que nous nous identifions avec elles.) À moins de tenir compte des processus d'incorporation et d'identification de ce genre, la moralité (et même la prudence) apparaîtra tout à fait mystérieuse, mais je ne me prononcerai pas davantage ici sur cette question. Pour quelques remarques préliminaires à ce sujet, cf. la section IX de mon article « Alienation, Consequentialism, and the Demands of Morality », *Philosophy and Public Affairs*, vol. 13, 1984, p. 134-171.

- ³⁶Ici, tout comme ailleurs, nous entendons par « théorie morale contemporaine » les idées dominantes au sein des académies, et nous désignons par « discours moral contemporain » les pratiques répandues de l'argumentation publique en matière de moralité, généralement opérantes dans le cadre des sociétés où les sciences empiriques ont connu les plus hauts niveaux de développement. Rappelons de nouveau que le réalisme moral, tout comme le réalisme scientifique, ne s'engage pas à admettre l'idée d'un consensus global.
- ³⁷Toutefois, de telles interdictions ont historiquement témoigné de limitations dans leur portée qui ne sont plus reconnues aujourd'hui comme valides. La tendance contre de telles limitations représente un exemple de la première tendance [*pattern*] évoqué ci-dessus, à savoir la tendance vers une généralité accrue.
- ³⁸Qui plus est, les explications présentent un caractère très général, et opèrent à un niveau de description qui n'est pas à même de discriminer entre des hypothèses qui reposent sur la théorie particulière de ce qui est moralement bien [*rightness*], proposée ici, ainsi que d'autres théories qui se rapprochent plutôt de celle-ci. (Grosso modo, ceux qui caractérisent ce qui est bien [*rightness*] en termes de rationalité instrumentale, relative au bien non moral de celles et ceux concernés, mais qui diffèrent sur les détails de cette rationalité instrumentale – par exemple, est-elle simplement maximisante ou partiellement distributive? – ou du bien [*goodness*] non moral – par exemple, est-il réductible au plaisir? Pour une discussion de conceptions concurrentes et quelque peu divergentes, voir la section VI). Si on ne sert pas de la méthode employée ici pour faire des choix parmi des conceptions concurrentes et similaires, l'analyse empirique doit être alors beaucoup plus fine [*fine-grained*]. Des remarques similaires s'appliquent aux théories faibles et étendues de la rationalité individuelle, décrites ci-dessus.
- ³⁹Il importe peu que nous affirmions qu'il n'ait *aucune* raison ou qu'il y a simplement une raison infiniment faible de poser cette action. Je suppose que nous pourrions même affirmer qu'une personne a une raison infiniment faible de faire quoi que ce soit – même de consacrer une quantité énorme d'efforts à purger des contradictions de ses croyances ou à purger des alligators d'un marécage lointain – qui pourrait de manière *concevable* s'avérer être à son avantage. Mais nous n'aurions alors aucune difficulté à garantir l'existence de raisons infiniment petites pour motiver la conduite morale. Cela permettrait, après tout, à une conception naturaliste du bien moral [*naturalized moral rightness*] de satisfaire à la thèse humienne de la practicalité, mais cette thèse ne serait pas pour autant dérobée de son intérêt.
- ⁴⁰François (duc de) la Rochefoucauld, *Réflexions, ou sentences et maximes morales, suivies de réflexions diverses*, Éd. Jean Lafond, Paris, Gallimard, 1976, p. 79. La Rochefoucauld a apparemment emprunté cette phrase du clerc, Du Moulin. Je suis reconnaissant envers Barrington Moore Jr. de m'avoir rappelé cette citation. Cf. son ouvrage *Injustice*, p. 508.
- ⁴¹Contrastons ces idées avec le relativisme présenté par Harman au sujet des obligations [*ought*] dans *The Nature of Morality*. Harman adopte la première des deux stratégies présentées ci-dessus, de sorte qu'il conserve la relation entre les obligations morales d'un individu et la raison (instrumentale) qui motive sa conduite. Il défend son approche, en partie, en faisant valoir que si nous supposions qu'Hitler était engagé dans la poursuite rationnelle de ses fins, un jugement « interne » tel que « Hitler n'aurait pas dû (moralement) avoir tué 6 millions de juifs » serait « faible » ou « étrange » en comparaison avec un jugement « externe » tel que « Hitler était quelqu'un de mauvais [*evil*] » (voir, p. 107 sq.). Je serais porté à penser le contraire, à savoir qu'il est trop « faible » et « étrange » de formuler une théorie morale selon laquelle Hitler serait jugé comme étant quelqu'un de foncièrement mauvais (ce qu'Harman prétend accomplir avec son type de relativisme, sans présenter la moindre explication), mais dans laquelle 'Hitler (moralement) n'aurait pas dû agir comme il l'a fait' est faux.
- ⁴²Ce besoin ressenti se reflète également dans la codification des lois et dans le développement de théories légales. Bien que la loi puisse sembler parfois artificielle, les conditions sociales générales et les besoins qui ont motivé son développement sont pourtant bien réels. En effet, l'artifice élaboré de la loi et de son langage reflète en partie la nécessité pressante de surmonter le niveau de description du sens commun pré-théorique.

⁴³Je suis redevable à plusieurs personnes, notamment Peter Achinstein, Robert Audi, Annette Baier, Michael Bratman, Stephen Darwall, Allan Gibbard, Thomas Nagel, Samuel Scheffler, Rebecca Scott, Nicholas Sturgeon, Nicholas White, et les éditeurs de *The Philosophical Review*, qui ont eu la gentillesse de me faire part de leurs commentaires sur des versions antérieures ou des présentations de cet article.

ÉTUDE CRITIQUE

RÉPONSE À UNE CRITIQUE

FRANCESCO PAOLO ADORNO

PROFESSEUR DE PHILOSOPHIE MORALE, UNIVERSITÉ DE SALERNE

RÉPLIQUE À ALICE LANCELLE ET MARJOLAINE DESCHÊNES,
« FRANCESCO PAOLO ADORNO, FAUT-IL SE SOUCIER DU CARE?.
UNE ETUDE CRITIQUE », *LES ATELIERS DE L'ÉTHIQUE/THE ETHICS FORUM*,
vol. 10, no. 3, 2015, p. 168-190.

Quand j'ai commencé à rédiger mon ouvrage sur le *care*, je savais que j'allais m'exposer à des critiques – ce qui est absolument normal et même salutaire. Toutefois, j'avais espéré que l'éventuelle discussion que mon travail pouvait engendrer fasse l'économie des accusations de pensée néolibérale, antiféministe et par conséquent réactionnaire, et que l'on discute les thèses de fond. Pour cela, il aurait fallu comprendre que ma critique des éthiques du *care* n'était que la manifestation d'une préoccupation réelle concernant leur validité et leur force théorique.

Car mon livre ne discute ni du féminisme, ni des femmes, mais du *care*. La différence est substantielle. Certes, les théories du *care* sont nées et se sont développées grâce à des femmes qui avaient à cœur la situation des femmes dans la société actuelle (cela dit, je ne crois pas que les discriminations que les femmes subissent n'intéressent qu'elles), mais ce souci premier a été élargi à un problème, la place du *care* dans nos sociétés, qui dépasse les questions de sexe ou de genre. Les auteurs que j'ai discutés, à partir de Joan C. Tronto qui a été la plus lue, la plus commentée et la plus utilisée en France¹, ne sont pas parmi ceux ou celles qui font du *care* une question de genre, qui par ailleurs n'a jamais été traitée en tant que telle dans mon livre. Mon parti pris a été de considérer que le *care* est une théorie éthique qui a des retombées pratiques sur l'ensemble de la société. Il ne me semble pas inutile de remarquer qu'une théorie est bonne ou mauvaise indépendamment du sexe de ses théoriciens.

Tout en trouvant particulièrement injuste cette réduction du *care* au féminisme, un biais que j'ai essayé de ne pas reproduire dans mon ouvrage, j'en comprends la raison stratégique. Affirmer l'équivalence totale entre féminisme et *care* permet de taxer d'antiféminisme tous les hommes qui s'aviseraient d'énoncer des critiques à l'encontre du *care*. Ainsi, il paraît que mon livre est entaché de « biais de genre », et que j'ai essayé rien de moins que de « discréditer les mouvements féministes et leur pertinence ».² Comme l'affirment Alice Lancelle et Marjolaine Deschênes, ma critique d'une théorie « féministe » est le signe d'un discours antiféministe qui voudrait « instruire les femmes, de son point de vue philosophique mâle et dominant ». On peut se demander au passage si les critiques acceptables du *care* ne pourraient venir que des femmes. Le cas échéant, celles qui se voient critiquées auraient un argument tout trouvé pour les récuser d'avance avant même d'en avoir entendu le contenu. J'avais essayé d'anticiper cette possible incompréhension dans mon introduction :

le chercheur (*a fortiori* si c'est un homme) qui s'aventurerait à émettre des doutes sur la consistance de cette théorie doit s'attendre à être traité de « macho » (ce qui finalement pourrait ne pas être l'insulte que l'on croit), d'intellectuel de seconde zone tirant à boulets rouges sur une théorie qui n'est pas hégémonique et donc à sa portée, ou encore de mauvais lecteur incapable de voir sa radicalité et sa nouveauté.³

Si elles m'avaient lu avec une plus grande attention, au lieu de gaspiller leur énergie à débiter des accusations sans fondement, mes critiques auraient pu se

concentrer sur des points plus importants de mon travail. D'autant plus qu'elles reconnaissent qu'au fond mes « critiques et remarques sont souvent pertinentes », ce qui les conduit à déplorer que je « ne prenne pas le projet au sérieux ni ne tente de résoudre les problèmes [que j'] expose avec plus de générosité et de finesse ». (p. 184) Encore que je comprenne ce que signifie résoudre des problèmes avec générosité et que l'on puisse considérer cette critique d'un texte qui se veut théorique comme pertinente, il apparaît que ma « thèse principale reste que les éthiques du *care* n'ont pas les moyens de renverser les rapports de pouvoir qu'elles dénoncent et risquent même d'empirer ce qu'elles critiquent », ce qui vaut reconnaissance à demi-mot du fait que je ne suis peut-être pas si antiféministe que mes critiques veulent le croire. Donc, je tiens à rassurer mes lectrices éventuelles : non, je ne pense pas que le féminisme n'arrivera jamais à transformer le monde, je ne crois pas qu'il soit nuisible pour les femmes et surtout je n'ai jamais soutenu « à l'insu de mon plein gré » la thèse soutenant que le féminisme met en péril l'ordre (p. 186), sans qu'elles précisent de quel ordre il s'agit.

Je comprends que le ton ironique et provocateur de mon travail, que j'assume totalement, ait contribué à faire passer au second plan le fond de ma contribution à la discussion des éthiques du *care*, mais la lecture des ouvrages sur le *care* ne peut pas laisser indifférent : les exemples donnés par Sarah Ruddick, Nel Noddings, Joan C. Tronto, et j'en passe, pour étayer leurs thèses sont assez pittoresques et méritent d'être repris et commentés pour en montrer les tenants et aboutissants. Il n'est pas surprenant que, après la lecture de quelques centaines de pages qui se veulent fondatrices de rien de moins qu'une nouvelle éthique, qui ont eu et continuent d'avoir un énorme retentissement, et dans lesquelles on lit toutefois des choses que l'on trouverait déplacées dans un mémoire de maîtrise, on se laisse emporter par un mouvement d'humeur. Les sentiments, les émotions et les passions que suscite une lecture doivent toutefois être verbalisés et justifiés, faute de quoi ils ne seraient rien de plus que l'expression d'une idiosyncrasie : c'est dans cet esprit que j'ai voulu transformer ma déception, ma préoccupation et parfois aussi mon énervement, face à des ouvrages qui ne me semblaient pas à la hauteur de leur réputation, en une critique raisonnée.

En effet, il me semble qu'une des tâches qui incombe à tout philosophe qui se trouve aux prises avec une théorie est justement d'en comprendre la structure profonde, de valider les définitions qu'elle propose, d'en chercher la cohérence et éventuellement d'indiquer les raisons d'une évaluation négative et, si parti pris il y a, d'en justifier l'existence. Ce que l'on demande au minimum à une théorie morale est qu'elle soit justifiable, cohérente et normative : j'ai essayé de montrer que sur ces trois points les éthiques du *care* sont plutôt insatisfaisantes et j'aurais souhaité que l'on prenne en compte ces critiques théoriques plutôt que d'intenter un procès à mon ouvrage en m'accusant d'éprouver de la haine pour les femmes. Mon souhait implicite a été totalement déçu : au lieu de discuter la lettre et l'esprit de mes critiques, Lancelle et Deschênes ont préféré se lancer dans une sorte de réquisitoire « résolument féministe », disent-elles, qui frôle parfois l'insulte, contre mon interprétation des éthiques du *care*.

Je reprendrai donc certaines analyses formulées dans mon livre, moins pour critiquer mes critiques que pour montrer encore une fois en quoi les théories du *care* sont insatisfaisantes. J'éviterai de relever les imprécisions, les passages confus voire les erreurs qui émaillent la critique de ma critique et qui concernent les auteurs qui sont au centre du *care*, de peur de me faire taxer une nouvelle fois de mâle qui veut faire la leçon aux femmes, pour me concentrer en revanche sur des points théoriques importants qui méritent d'être discutés, en partie parce que l'on m'attribue des thèses que je n'ai jamais soutenues et en partie parce que la manière dont ils sont repris dans cet article démontre au moins une certaine incompréhension des finalités et des articulations internes de mon travail qui, je le rappelle, ne voulait être rien de plus, mais rien de moins non plus, qu'une introduction aux éthiques du *care*.

Avant de reprendre les critiques adressées au contenu de mon ouvrage, je voudrais d'abord m'arrêter sur les lacunes et les absences qui, semble-t-il, le rendent incomplet et donc infidèle. Cette accusation naît à mon sens du fait que mes lectrices n'ont pas vu que mon but était de faire une présentation critique des textes fondateurs de cette théorie à partir d'un point de vue strictement philosophique et qui, de manière encore plus stricte, se proposait de discuter la prétention du *care* d'être une théorie éthique. Pour le dire clairement, Deschênes et Lancelle n'ont même pas soupçonné que mon livre était un livre de philosophie morale qui s'occupe avant tout du rapport « incestueux » entre philosophie morale et psychologie cognitive dans les éthiques du *care*. Ce problème a été totalement ignoré dans leur étude alors qu'une lecture plus attentive n'aurait pu manquer cet aspect fondamental qui justifie un grand nombre de critiques que j'adresse au *care*. La perspective choisie m'a évidemment mené à faire des choix, tant dans la sélection des auteurs discutés que dans les problèmes pris en compte.

Cette nécessité de faire des choix et de limiter le champ d'analyse, que chaque auteur de livre connaît bien, a été ignorée à priori, car mes critiques affirment sans nuances que : a) je fais semblant de ne pas connaître les travaux fondamentaux (!) de Guillaume Leblanc et de Marie Garrau, ce qui m'a amené à donner une « présentation du concept de vulnérabilité... erronée à plusieurs titres » (p. 179), mon « erreur », qui invalide une bonne quinzaine de pages de mon travail (Adorno, 2015, pp. 100-118), se résumant finalement à ne pas avoir tenu compte des thèses de ces auteurs. Sans vouloir diminuer la portée des réflexions de Leblanc et de Garrau, je ne les ai pas mentionnés simplement parce que leurs analyses, pour intéressantes qu'elles soient, n'apportaient strictement rien à ma perspective, à la différence des suggestions qui me venaient de Jacques Lacan, Martin Heidegger, Hannah Arendt, Charles Taylor, Michel Foucault, etc.; b) je ferais mieux de prendre connaissance de « récentes avancées américaines », pour combler l'étroitesse de mes références; c) mon travail présente des lacunes manifestes, d'autant plus graves qu'elles se trouvent « dans un ouvrage prétendant introduire aux théories du *care* » (p. 180) parce que je ne discute pas les thèses de Susan Moller Okin et d'Eva Feder Kittay sur la théorie de la justice de John Rawls.

Mis à part le fait qu'une introduction à un problème n'est pas une encyclopédie, et que j'ai beaucoup plus de sympathie pour des textes français ou américains classiques que pour d'obscures quoique « récentes avancées américaines », les raisons de la présence ou de l'absence de certains théoriciens du *care* dans mon ouvrage sont parfaitement claires. Je me suis attardé sur des auteurs qui ont structuré le débat sur le *care* et se sont concentrés sur le *care* de manière en même temps spécifique et générale : que leurs thèses ne soient pas ou peu d'actualité (ce qui reste quand même à prouver) ne veut pas dire qu'il ne faut plus en parler ou qu'il aurait fallu parler d'autres ouvrages. Si par ailleurs le problème est que la plupart des auteurs cités ne défendent pas une conception « féministe » du *care*, un choix dicté par le fait que, je le répète, je n'étais et ne suis toujours aucunement intéressé par les questions de genre, il s'agit d'un parti pris de la part de mes critiques qui ne me concerne pas. Comme Martha C. Nussbaum l'a remarqué, il est absolument vrai que le libéralisme de Rawls n'arrive pas à discuter de façon satisfaisante les questions liées aux soins à apporter dans les moments de la vie où on se trouve dans des situations de dépendance et au rôle politique de la famille, et notamment aux problèmes de justice dans la famille, mais mon livre voulait dresser un bilan critique des éthiques du *care* en tant que théories, au-delà de toutes les questions de genre.

Ce manque d'intérêt pour les questions de genre explique l'absence d'Okin dans mon livre. Son problème est celui de la division injuste du travail dans la famille, donc les rapports de genre dans la famille, et elle plaide pour « un partage égal des responsabilités familiales entre les sexes particulièrement concernant le soin des enfants ». ⁴ Son ouvrage, qui mérite sans aucun doute toute l'attention qu'il a suscitée, discute de problèmes qui sont à mon avis secondaires dans les théories du *care*. Que les critiques qu'elle adresse, comme tant d'autres, à Rawls soient valables ou non, il n'y avait pas de raison de les mentionner dans un ouvrage consacré au *care* et non aux rapports de genres ni à la théorie de la justice.

Les raisons justifiant l'absence de l'ouvrage de Kittay sont différentes et plus complexes. Même si mes critiques affirment, sans plus, qu'elles comptent parmi les plus fécondes dans l'univers féministe (s'agissant de critiques féministes, jouiraient-elles d'un statut spécifique?), j'ai du mal à voir en quoi ses critiques à Rawls pouvaient apporter quelque chose de plus et de différent de celles que j'ai citées et discutées. De plus, sa confrontation avec la théorie de Rawls est opérée d'un point de vue très spécifique qui consiste à conduire une étude sociologique et politique sur la situation des *dependency workers*. À l'intérieur de cette catégorie, Kittay pointe les difficultés rencontrées par les *dependency workers* qui s'occupent d'une catégorie spécifique de personnes, en l'occurrence d'individus lourdement handicapés comme sa fille Sessa. Je dirais donc que, chez Kittay, il n'y a pas une éthique du *care* de valeur universelle, mais la volonté de présenter les difficultés et les injustices qui touchent une part importante de la population, en sa plus grande partie constituée de femmes : les travailleurs de la dépendance. Tout son travail me semble par ailleurs une tentative de montrer que, ne considérant pas les nécessités de cette catégorie de personnes,

la théorie de la justice arrive à des résultats insatisfaisants, ce qui l'amène à plaider pour une meilleure justice sociale et à proposer d'intégrer les deux principes de justice en un troisième qui dit : « to each according to his or her need for care, from each according to his or her capacity for care ; and such support from social institutions as to make available resources and opportunities to those providing care, so that all will be adequately attendend in relations that are sustaining ».⁵

Au fond, Kittay, comme Nussbaum d'ailleurs, se propose de corriger la théorie de la justice de Rawls et non pas de la remplacer par une éthique ou une politique du *care*, comme le fait en revanche une théoricienne comme Tronto. Elle poursuit ainsi le même but que la théorie de la justice de Rawls : faire en sorte que dans une situation de rareté des ressources qui conduit à des conflits inévitables dans leur distribution, ceux-ci puissent être réduits pour permettre à chacun de réaliser son idée de vie bonne. Ainsi, elle indique que son idée d'égalité engendre des revendications qui doivent permettre d'améliorer la situation des *dependency workers*, car elle se propose de définir les modalités suivant lesquelles « one can give care without the caregiving becoming a liability to one's own well-being ».⁶ S'occuper des autres est un lourd fardeau et il n'est pas juste que les femmes soient abandonnées à elles-mêmes, ce sur quoi on ne peut qu'être d'accord; mais je ne vois pas en quoi, dans son ouvrage, il est question d'une éthique du *care*.

J'en viens maintenant à ce qui, dans mon livre, a été peu et mal compris. Je suis parti du constat que, d'une manière générale, une des raisons principales de la faiblesse des théories du *care* est à rechercher dans leur statut plutôt incertain qui oscille, suivant les auteures, entre philosophie, psychologie, pédagogie et/ou sociologie. Ce mélange des savoirs est opéré au nom de la transdisciplinarité ou de l'interdisciplinarité nécessaire et souhaitable pour tout discours qui se veut innovant. Une des conditions fondamentales des innovations théoriques est, semble-t-il, leur capacité de briser les frontières entre savoirs, de ne pas se plier aux lois du genre, mais de les construire. Comme on le sait, l'interdisciplinarité est devenue une nécessité quand, à la fin des années 1950, ayant constaté l'existence d'une rupture profonde entre sciences humaines et sciences dures, entre deux cultures, littéraire et scientifique⁷, on a compris que la complexité croissante de notre monde demandait de construire des passerelles entre ces deux formes de savoir. L'interdisciplinarité est donc justifiée par la nécessité de solliciter des connaissances venant de domaines très différents à des fins spécifiques de compréhension et de connaissance de phénomènes complexes qui ne peuvent être appréhendés d'un seul point de vue. En l'absence de cette motivation, l'hybridation entre disciplines différentes, dans le meilleur des cas, ne produit aucun résultat valable et, au pire, autorise tout le monde à parler de tout et de n'importe quoi sans avoir aucune connaissance, avec des résultats qui sous des airs de grande compétence ne sont souvent que du pur et simple baratin.

En signalant l'existence de cette pratique dans les textes sur le *care*, il ne s'agissait nullement pour moi de chercher « à tout prix des incongruités épistémolo-

giques et théoriques dans les travaux de Gilligan » (p. 173) et des autres théoriciens mentionnés, comme on me le reproche, mais de considérer que chaque champ théorique possède une technicité propre qu'il faut connaître et comprendre quand on se propose de le « contaminer » avec d'autres savoirs. Cette méconnaissance conduit à des malentendus d'une grande portée, comme j'ai essayé de le montrer à propos de la différence, sur laquelle les éthiques du *care* font l'impasse, entre la question métaéthique de la motivation morale et la question de la justification des principes qui nous permettent d'identifier une action comme bonne ou mauvaise, juste ou injuste. De plus, en parlant du *care* comme si c'était une théorie qui avait toujours le même objet et mobilisait toujours les mêmes savoirs, on risque d'ignorer les différences entre ses théoriciens, mais aussi de passer sous silence les différences entre les perspectives à partir desquelles on discute du *care*. Ainsi, Carol Gilligan est une psychologue dont tout le travail est centré sur la volonté de montrer que les femmes se développent moralement en suivant des chemins différents de ceux des hommes, car elles ont une perspective différente sur autrui. Mais peut-on dire que les soucis théoriques de Kittay ou de Dietmar Bubeck sont les mêmes et que par conséquent leurs réflexions sur le *care* poursuivent le même but? Il me semble que dans leurs deux ouvrages, on trouve plutôt des réflexions sociopolitiques sur la manière dont deux formes de *care* ont fonctionné comme outil de domination des hommes sur les femmes. Ou encore, mis à part le terme, le travail de Sandra Laugier a-t-il quelque chose à voir avec le maternalisme d'une Sarah Ruddick ou d'une Nel Noddings? Il me semble que rendre chaque ouvrage sur le *care* à son lieu d'origine avait pour fonction de clarifier la situation générale et de montrer la spécificité de chaque réflexion, tout comme, en resituant dans leur perspective théorique originale les différents travaux sur le *care*, on se donnait la possibilité de comprendre la faiblesse et la force de chaque auteur et de son travail.

Mais il y a une raison supplémentaire à l'attention que j'ai portée aux mélanges des savoirs dans les théories du *care*, que j'expliquerai en reprenant et en expliquant mieux les remarques sur la rhétorique phonétique dans l'ouvrage de Gilligan.

À la fin d'un long passage (p. 173) qui commence en affirmant avec fermeté qu'en retraçant les origines du *care* j'avais le but « de montrer que Gilligan n'est pas si originale qu'on le prétend, puisque Ruddick et Noddings la précèdent » (alors qu'il s'agissait simplement de montrer qu'au début des années 1980 existait un réel intérêt pour ce concept : outre Gilligan, Ruddick et Noddings s'occupaient du *care* de manière continue),⁸ je suis sommé de m'expliquer sur mon but (supposé, car nulle part j'ai soutenu une telle thèse) qui consiste à montrer « que la voix des femmes qu'elle [Gilligan] veut faire entendre » n'est pas une tâche très originale. Sur ce point, mon

argument consiste à dire que, mettant l'accent sur la voix des personnes (des sujets) à la défaveur de la voix hégémonique (normes), Gilligan reconduit une « métaphysique qui ne dit pas son nom », soit une métaphysique de la présence telle que Jacques Derrida l'a pourtant déjà si bien critiquée (p. 35). Argument lui-même glissant s'il en est : tandis que Gilli-

gan souhaite en 1982 faire entendre les voix des filles et des femmes dans un champ de recherche qui jusqu'alors ne s'est effectivement intéressé qu'aux garçons et aux hommes, Adorno ne voit là rien que la répétition d'une métaphysique continentale. On ne pourrait manquer plus radicalement la portée des travaux féministes de Gilligan – féminisme qu'Adorno redoute et dénonce tout au long de ce livre, ce qui, selon nous, est plus proprement réactionnaire que « réactif ». [...] Nous interrogeons ici la pertinence d'en appeler à ces considérations phénoménologiques alors que la discussion porte chez Gilligan sur le développement moral des enfants. D'un point de vue à la fois existentiel, clinique, éthique, moral et politique, le combat de Derrida au sujet de la phénoménologie pure husserlienne et des déconstructionnistes de son école paraît tout à fait déconnecté de la problématique. Notons par ailleurs que Sophie Bourgault a soutenu de façon convaincante que Gilligan, Noddings et Ruddick, plutôt que de reconduire une métaphysique de la voix et de la parole (comme Adorno le suppose pour Gilligan, bien que Bourgault ne le cite pas), introduisent au contraire une éthique de l'écoute qui tranche tout à fait avec, justement, le manque d'écoute dont fait preuve la tradition logocentrique occidentale... Bourgault mentionne que les auteures de l'écoute (du *care*) trouvent écho chez Fiona Robinson, spécialiste des théories politiques et des relations internationales, qui développe une éthique des relations internationales axée sur le dialogue et l'écoute d'autrui (pp. 173-174).

Je passe sur le fait que les deux tiers du travail de Gilligan sont consacrés aux dilemmes qui se présentent à de jeunes femmes, dont l'interprétation des réponses s'intercale avec des analyses des travaux de Sigmund Freud, d'Erick Erikson, ou avec des œuvres littéraires et cinématographiques, et non aux enfants; que Bourgault ne pouvait pas me citer parce que mon ouvrage a été publié après son article et que finalement, comme je l'ai dit et répété, je considère les éthiques du *care* plus inconsistantes et inutiles – en tant que théories – que menaçantes, pour en venir à la question fondamentale.

Le contresens touchant à mon usage de la critique de Derrida de la phénoménologie husserlienne ne pouvait être plus complet : j'aurais démontré de cette manière rien de moins que ma totale incompréhension du travail de Gilligan. Comme mes critiques m'y invitent (« Si donc Adorno souhaite tout de même soutenir qu'une métaphysique du sujet (asexué) est reconduite par Gilligan à son insu, nous attendons l'argumentaire qui viendrait un peu valider cette hypothèse aussi rapide que contradictoire » (pp. 173-174).) je vais étayer mon propos sur ce problème, parce que je n'ai rien soutenu de semblable.

En premier lieu, il faut noter l'insistance avec laquelle la question de la voix est présente dans les théories du *care* en général et pas seulement chez Gilligan : il n'y a pratiquement pas un texte qui ne paye son tribut à cette rhétorique phonétique. De Kittay à Tronto, de Held à Okin, de Robinson à Baier l'insistance sur la voix réprimée, censurée, ignorée des femmes est commune. Et ce qui est commun est aussi l'idée que cette voix n'est qu'une métaphore de la subjectivité

des femmes. Mais alors, pourquoi autant d'intérêt pour la voix? Pourquoi utiliser cette image, qui devient une figure rhétorique (métonymie ou synecdoque), plutôt qu'un terme beaucoup plus courant et beaucoup plus clair : sujet ou subjectivité? Quelle signification le terme « voix » véhicule-t-il que le terme « subjectivité » ne reprendrait pas?

J'ai en effet souligné que la question était que « l'appel à l'émergence d'une autre voix nous renvoie à l'espèce de statut privilégié, assez ambigu, que l'on accorde à la voix » et, pour être encore plus précis, j'avais insisté sur le fait que « la question n'est pas de préférer une voix à une autre; c'est celle du statut privilégié de la voix en général ».⁹ La réponse a été justement de cerner cet élément différentiel qui fait sens dans la voix et qui est absent dans la notion de subjectivité. Or la voix exprime quelque chose, elle est surtout un moyen d'expression. Et ce que la voix exprime est une intériorité qui est supposée faire corps avec la voix qui l'exprime : pas de restes, pas de déchets, pas de différences, serait-on tenté de dire, entre le vouloir dire de la voix et celle ou celui qui veut dire et exprime ce qu'il veut dire à travers sa voix. La nature de la femme se manifeste dans sa voix qui l'exprime, ce qui a la fâcheuse conséquence de relier de manière exclusive ce qui est dit, le vouloir dire à la voix qui l'exprime. Pour le dire de manière plus claire : la voix est naturelle, la subjectivité est culturelle. Ce qui manque à la subjectivité est la présence d'un élément naturel irréductible : la voix est toujours (au moins) sexuée et en tant que telle immédiate et spontanée. La subjectivité est construite et donc soumise à une loi considérée étrangère et qui considère étrangères les différences naturelles. Dans la voix se manifestent les puissances de la nature que la loi ne peut maîtriser. Au fond, cette rhétorique de la voix nous amène à retrouver l'opposition fantasmatique entre Déméter et Abraham qui semble bien agir en arrière-plan des textes sur le *care* et au-delà de toute dénégation d'une opposition tranchée entre masculin et féminin, entre une morale masculine et une morale féminine. C'est justement ce que reconnaît Gilligan elle-même qui, tout en admettant une certaine résistance de la part des psychologues à utiliser ce terme, affirme l'avoir préféré à « soi » car, dit-elle, il est plus précis et moins abstrait : « à la fois corps et langage, la voix a aussi l'avantage d'ancrer la psychologie dans la biologie et la culture sans non plus se réduire à l'une ou à l'autre ».¹⁰ De plus, la *phoné* réitère la structure de la métaphysique qui consiste à affirmer que la voix est un moyen d'expression de ce que l'on ne voit pas à travers et dans ce qui est physiquement présent.

On ne comprendra pas au fond cette insistance sur la voix en tant que symptôme métaphysique si on fait l'impasse sur l'analyse que développe Derrida du rapport entre voix et écriture chez Edmund Husserl et Martin Heidegger.¹¹ La condamnation de l'écriture chez Heidegger donne un relief théorique au sort qui est fait à la voix : alors que la première éloigne de la nature et pour cela est soumise à une critique impitoyable de la part de Heidegger, Derrida relève le privilège que le philosophe allemand accorde à la voix, de manière que le phonocentrisme n'est qu'une figure du logocentrisme, que les théories du *care* ont pourtant en ligne de mire. Ce qui m'a amené à écrire que cette insistance sur la voix au fond est contradictoire, car elle réitère le mécanisme même auquel les théories du

care veulent se soustraire. On me dira que, même si j'ai raison sur ce point, la question n'est pas là, mais dans le fait que les femmes n'arrivent pas à se faire entendre, qu'elles réclament toujours la place accordée à leurs pensées, qu'on la leur refuse, et que, pour la première fois, Gilligan a porté au-devant de la scène cette spécificité. Si le diagnostic historique ne souffre pas de critique, en mettant l'accent sur la voix Gilligan fait quelque chose d'autre. Elle tente de définir une morale, sinon deux, fondées sur la nature humaine, sur la psychologie plus précisément, ce qui, comme j'ai essayé de le démontrer grâce à la référence à Derrida, est en tout et pour tout une reconduction de la métaphysique classique.

Mais c'est aussi le symptôme d'une difficulté majeure des éthiques du *care*, qu'elles partagent avec un certain nombre de théories en philosophie morale : l'idée qu'il est possible de fonder une morale sur des présupposés psychologiques. En philosophie morale, on assiste depuis au moins une trentaine d'années au développement de formes de naturalisme moral qui prennent souvent l'aspect de théories de psychologie morale. Avec l'ouvrage de Gilligan et l'insistance sur l'importance des sentiments et des émotions en morale, les éthiques du *care* reprennent et articulent les arguments centraux de ce courant plutôt varié qui était justement l'aspect que je souhaitais mettre en évidence dans le chapitre « Morale des sentiments, morale de la raison ».

L'affirmation que tous les ouvrages sur le *care* sont structurés autour de l'opposition entre éthique des sentiments et morale de la raison me semble difficilement contestable : ce n'est certainement pas moi qui m'efforce de les structurer de cette manière, comme Lancelle et Deschênes le soutiennent (p. 176). Dès lors, je crois qu'il était tout à fait légitime d'analyser cette opposition pour en tester la validité et la normativité en se demandant s'il existe une morale normative fondée sur les sentiments et si elle est en mesure d'éliminer les défauts qui grèvent une morale fondée sur des concepts plus traditionnels comme la justice. Il m'a donc semblé nécessaire d'évaluer la capacité normative des concepts d'une éthique du *care* qui prétend se fonder sur les sentiments. Parce que, qu'il s'agisse d'une morale conséquentialiste, d'une éthique des vertus ou d'une morale déontologique, une théorie morale est définie par sa capacité à indiquer ce qu'il faut faire pour être bons et/ou justes, à nous expliquer les motifs de cette action et à justifier ces indications de comportement.

Sans entrer dans des détails trop techniques qui concernent l'ontologie des valeurs morales, leur épistémologie ou encore le sens des concepts moraux, je me suis proposé d'abord de voir quels concepts normatifs étaient proposés et j'ai voulu ensuite tester leur cohérence par rapport à une question de métaéthique qui est très souvent ignorée, par incompetence ou par négligence, et qui concerne la différence nécessaire entre la définition des motivations d'une action morale et la recherche des justifications de cette action et de ses principes. Pour être plus clair : à travers la définition des motivations morales, on essaye d'éliminer la possibilité ou la présence de l'amoraliste, de celui qui ne veut pas se comporter moralement (ni immoralement par ailleurs); la définition des justifications de principes moraux tente de réduire ou d'éliminer les conflits entre

visions morales concurrentes parce que l'on recherche les raisons de ces actions. La première est une question de psychologie et la seconde une question purement philosophique. L'analyse des sentiments, des émotions, des passions dit en gros quels sont les motifs qui me poussent à agir (ou à ne pas agir) moralement, sans que l'action morale soit définie dans sa normativité, alors que l'analyse de mes raisons d'agir m'offre une justification de ce que je fais pour avoir un comportement moral. Or, les textes consacrés à l'éthique du *care* font l'impasse, à de rares exceptions près, sur cette différence, de telle manière que l'on considère que les sentiments peuvent jouer un rôle normatif. Quand on se demande quelles sont ces normes que les sentiments, tels que l'amour, produisent, la confusion devient si grande qu'on aurait le plus grand mal à en déduire des principes normatifs. Car que veut dire que les sentiments sont normatifs sinon que, en me confiant à mes sentiments, je peux me comporter bien ou de manière juste? Mais quand on se demande comment se réalise cette déduction, les réponses que l'on obtient sont plutôt décevantes.

Venons-en à un deuxième point tout aussi important : la critique du libéralisme par le *care*, ou plus précisément les rapports entre libéralisme et *care*. Il s'agit d'un topos qui est développé dans tous les textes sur le *care*, suivant les mêmes coordonnées théoriques, à quelques détails près. Le libéralisme, nous dit-on, propose une image idéologique des individus qui ne correspond pas à la réalité de leur vie : ils sont décrits comme des atomes, préoccupés uniquement de promouvoir leur intérêt personnel. Ce qui les intéresse est au fond d'avoir raison, car ils sont sans relations et sans sentiments et ne se soucient guère de ceux et celles qui leur permettent d'avancer dans la vie. Ils ne savent pas qu'ils sont vulnérables et qu'ils peuvent à tout moment basculer dans des situations de dépendance. Pour le libéralisme tous les individus sont autonomes, libres et indépendants, ceux qui ne le sont pas sont exclus et confinés aux limites de la vie sociale, ignorés et exploités. À ce sujet, j'ai d'abord fait remarquer qu'il y avait plusieurs libéralismes, et non un seul, et que par ailleurs le libéralisme classique, notamment celui de John Locke, qui est considéré comme l'auteur du texte d'origine de l'individualisme possessif et de l'individualisme atomiste, mériterait une évaluation plus attentive capable de l'isoler dans la galaxie libérale et surtout de rendre à sa pensée toute sa richesse, quelque peu réduite par des critiques hâtives.

Ensuite, j'ai aussi fait remarquer que la critique du libéralisme que l'on trouve dans les textes sur le *care* était insuffisante d'un point de vue philosophique et qu'elle se fondait sur des généralités qui n'arrivent même pas à faire la distinction entre libéralisme économique, libéralisme politique et libéralisme moral. On peut être libéral en morale sans souscrire aux thèses économiques du libéralisme qui se traduisent en une économie où le marché est roi. On peut penser que chacun a le droit de chercher son bonheur personnel sans que personne ne l'en empêche (libéralisme moral) et à condition qu'il respecte la liberté des autres d'en faire autant, sans que cette thèse conduise à accepter inéluctablement l'économie de marché (libéralisme économique). Le libéralisme moral défend le droit de chacun de poursuivre son idéal de bonheur, quel qu'il soit, dans la limite du

respect du même droit pour les autres; le libéralisme économique fait du marché le lieu de vérité de l'action politique. Si ces deux formes de libéralisme sont évidemment compatibles, on peut parfaitement soutenir la première sans accepter la seconde, comme Rawls lui-même l'affirme. Je doute fort que les individus qui demandent le respect de leur droit au bonheur, quel qu'il soit, puissent être considérés comme des égoïstes, occupés seulement de leur bien-être et de leur succès personnel – sous-entendu économique – (encore que les deux choses coïncident), et que par conséquent ils puissent être réduits à l'image du *self-made-man* sans liens ni relations, comme le décrit le *care*.

Mais le problème est bien ailleurs, c'est-à-dire dans la dynamique entre relation à soi et relation à autrui : comment les structurer et surtout à quelle relation donner la priorité? À ce sujet Gilligan opère un distinguo important. Elle isole d'une part une forme de *care* qui est qualifiée de « féminin », dans lequel les femmes donnent la priorité aux autres et mettent leurs exigences au second plan. Cette forme d'attention est justement un dévouement surrogatoire que les femmes de l'enquête rapportée dans *In a Different Voice* commencent à interroger et à critiquer, mettant en avant une relation à soi qui les amène à choisir l'avortement plutôt qu'une vie de sacrifice et de soumission, ou en tout cas de difficulté à se réaliser. Elle dégage d'autre part une autre forme de *care*, féministe celui-ci, tout à fait différente : il faut d'abord avoir une relation authentique avec soi-même pour pouvoir ensuite trouver une relation avec les autres. La seconde est aussi précieuse que la première, mais elle vient après, car elle se fonde sur une bonne relation à soi.¹² Ce texte au fond ne présente rien de nouveau par rapport à ce qui était proposé dans *In a Different Voice*, là où Gilligan détaille sa vision de l'évolution morale : il y a d'abord une attention à soi-même, entièrement consacrée à la prise en charge de ses besoins pour pouvoir assurer sa survie, ensuite un *care* féminin et enfin un équilibre entre prendre soin de soi et prendre soin des autres.¹³ Mais Rawls dit-il autre chose? La question n'est-elle pas toujours la même? L'un et l'autre ne s'interrogent-ils pas sur la meilleure manière de réaliser ses aspirations, son plan de vie au-delà des intrusions des uns et des autres, ou pour le dire autrement, comment respecter les droits des individus à choisir librement leur conception du bien, réduisant au minimum les formes de domination et de gouvernement hétéronomes?

Quant au libéralisme économique qui s'est traduit dans l'idéologie du marché, j'ai émis l'hypothèse, suivant les suggestions de Foucault et d'Arendt, que le *care* était bien plus compatible que ce que ces théoriciens imaginent avec cette idéologie. Foucault a montré que la manière de gouverner les individus dans la modernité était absolument spécifique, parce qu'ancrée dans la vie biologique. La souveraineté médiévale s'exerçait grâce au prélèvement que le souverain pouvait à tout moment réaliser sur les biens de ses sujets : tout lui appartenait, même leur vie, et il pouvait en disposer comme bon lui semblait. En ce sens, il avait un droit de mort sur les individus, mais il s'intéressait peu à la manière dont ils vivaient. La souveraineté contemporaine est une forme de gouvernement – terme à prendre dans son sens étymologique – de la vie des individus. Il n'est plus question de prélever des biens, mais de faire en sorte que les indivi-

dus vivent le plus longtemps et le mieux possible. La satisfaction des besoins vitaux des individus est le levier que le pouvoir contemporain utilise pour gouverner. La gestion de la vie biologique des individus est devenue le point d'ancrage du pouvoir. La définition du *care* de Tronto et Fisher que je rappelle me semble tout à fait compatible avec cette forme de gouvernementalité moderne. La centralité du besoin et surtout du *devoir* de le satisfaire, avec son cortège de notions comme la responsabilité, l'attention aux autres, la *responsiveness*, semble bien être un symptôme de cette compatibilité. Notons que cette centralité du besoin revendiquée à tout moment par le *care* n'est pas suivie par une *théorie* des besoins, et souvent même pas par une définition des besoins que le *care* est censé satisfaire. Alors, plutôt que d'écarter avec agacement cette hypothèse, comme le fait par exemple Fabienne Brugère, ou de l'ignorer totalement comme le font Deschênes et Lancelle, ne vaudrait-il pas mieux essayer de l'analyser en profondeur? Entre le refus pur et simple des théories de la justice et l'accent mis sur le devoir de répondre sans limites aux besoins des autres, n'y a-t-il pas d'espace pour penser autrement le rapport entre justice et attention aux autres? Dans cette perspective, on pourrait aussi se demander pourquoi la dynamique du désir n'est jamais rappelée dans l'éthique du *care*? À titre d'hypothèse on pourrait penser qu'au fond cette absence n'est pas très surprenante : elle représente un symptôme de cette compatibilité implicite entre *care* et libéralisme.

Les lectrices et lecteurs qui ont eu la patience d'arriver jusqu'à la fin de ce texte auront peut-être remarqué la présence plutôt insistante, parfois même redondante, de termes comme « théorie », « éthique », « concept », utilisés à propos du *care*. J'ai souligné et souligne de nouveau cette présence pour une raison spécifique qui tient à la nature de mon travail sur le *care*. Mon but était tout à fait théorique : il s'agissait de tester les prétentions du *care* à être une théorie éthique, d'en mettre en évidence les incohérences, d'en indiquer les limites : en somme, un travail de critique philosophique qui prenait en ligne de mire des concepts et non pas des pratiques. Personne ne conteste l'importance du *care* dans la vie quotidienne : nous tous avons besoin que l'on s'occupe de nous, que des personnes, par amour ou par devoir, satisfassent nos besoins (encore faut-il être capable de les identifier correctement), que la vie soit pleine d'amour, d'attention, d'émotions et de sentiments en l'absence desquels, objectivement, elle serait bien moins attrayante. Toutefois, la tâche de la philosophie morale n'est pas de décrire ce que les gens font – la psychologie, la sociologie et toutes les sciences humaines s'en chargent –, mais d'indiquer ce que les gens devraient faire. Ce qui veut dire que l'évaluation pratique du *care* dans les hôpitaux, les écoles, les entreprises est du ressort de la sociologie, de la psychologie et des sciences humaines et non de la philosophie. De plus, il me semble que la fonction principale de la philosophie est critique et peut et doit se développer suivant deux impératifs : le premier nous vient de Friedrich Hegel, qui dit que « ce qui est bien connu est en général, pour cette raison qu'il est bien connu, non connu »;¹⁴ le second de Foucault, pour qui la tâche critique de la philosophie consistait dans sa capacité à rendre difficiles les gestes faciles. Les éthiques du *care* cumulent les deux défauts pointés par Hegel et Foucault. Elles sont faciles et compréhensibles, car elles mettent au centre de leur réflexion les bons senti-

ments, et elles ont un air de déjà-vu, ce qui les rend acceptables d'emblée : mon intention était de les décortiquer pour les rendre un peu moins évidentes qu'elles ne le paraissent.

Donc ma question était : les éthiques du *care* sont-elles de bonnes théories? Et la réponse était : non. Dans mon livre, je me suis efforcé d'expliquer les raisons de cette réponse, qui sont bonnes ou mauvaises, abstraction faite du sexe ou du genre de leur auteur, et qui tiennent surtout (mais non exclusivement) aux deux points que je viens de rappeler. Ai-je été entendu? La réponse est nette : non.

NOTES

- ¹ Voir Sherwin, Susan, *No Longer Patient. Feminists Ethics and Health Care*, Philadelphia, Temple UP, 1992.
- ² Lancelle, Alice et Marjolaine Deschênes, « Francesco Paolo Adorno, *Faut-il se soucier du care?* Une étude critique », *Les ateliers de l'éthique/The Ethics Forum*, vol. 10, no. 3, 2015, p. 168-190, p. 186. Dorénavant, ce texte sera cité simplement avec le numéro de page entre parenthèses.
- ³ Adorno, Francesco Paolo, *Faut-il se soucier du care?*, Paris, Édition de l'Olivier, 2015, p. 9.
- ⁴ Okin, Susan Moller, *Justice, genre, famille*, trad. fr. de Ludivine Thiaw-Po-Une, Paris, Flammarion, 2008, p. 24.
- ⁵ Kittay, Eva Feder, *Love's Labor. Essays on Women, Equality, and Dependency*, New York, Routledge, 1999, p. 113.
- ⁶ *Ibid.*, p. 66.
- ⁷ Percy Snow, Charles, *The Two Cultures*, Londres, Cambridge University Press, 2001 [1^{re} éd. 1959].
- ⁸ Ces travaux mêmes ont été précédés par d'autres, par exemple : Mayeroff, Milton, *On Caring*, New York, Harper Perennial, 1971.
- ⁹ Adorno, *Faut-il se soucier du care?*, p. 33.
- ¹⁰ Gilligan, Carol, « Résister à l'injustice : une éthique féministe du *care* », trad. fr. de Marlène Jouan, in Gilligan, Carol, Hochschild, Arlie & Joan C. Tronto, *Contre l'indifférence des privilégiés. À quoi sert le care*, édité et présenté par Pascale Molinier et Patricia Paperman, Paris, Payot, 2013, p. 40.
- ¹¹ Voir Derrida, Jacques, *La voix et le phénomène*, Paris, PUF, 1967; Derrida, Jacques, « L'oreille de Heidegger. Philopolémologie. (*Geschleth IV*) », in Derrida, Jacques, *Politiques de l'amitié*, Paris, Galilée, 1994, p. 341-419.
- ¹² Gilligan, « Résister à l'injustice : une éthique féministe du *care* », p. 43.
- ¹³ Gilligan, Carol, *Une voix différente. Pour une éthique du care*, trad. fr. d'Annick Kwiatek revue par Vanessa Nurock, Paris, Flammarion, 2008, p. 123.
- ¹⁴ Hegel, Georg Wilhelm Fredrich, *Phénoménologie de l'esprit*, trad. B. Bourgeois, préface, §31, Paris, Vrin, 2006, p. 79.