

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.

VOLUME 9 NUMÉRO 2 ÉTÉ/SUMMER 2014

4-28 Vers une éthique climatique plus efficace : motivations et incitations : **Michel Bourban**

29-49 Exploitation et obligation de travailler : **Pierre-Étienne Vandamme**

DOSSIER

Attitudes, valeurs et environnement

Sous la direction de **Antoine C. Dussault**

50-56 Introduction: **Antoine C. Dussault**

57-81 Thinking Globally, Acting Locally: Partiality, Preferences and Perspective: **Graham Oddie**

82-98 Sur la symétrie présumée entre valeurs et préférences: **Mauro Rossi**

99-109 Richness Theory: From Value to Action: **Gregory M. Mikkelsen**

110-129 Value Monism, Richness, And Environmental Ethics: **Chris Kelly**

130-148 Individualist Biocentrism vs. Holism Revisited: **Katie McShane**

DOSSIER

Métaéthique et métamétaphysique

Sous la direction de **Fabrice Correia** et **Christine Tappolet**

149-151 Plus on monte, plus on s'amuse: introduction: **Fabrice Correia** et **Christine Tappolet**

152-165 Sources, raisons et exigences: **Bruno Guindon**

166-189 Fitting-Attitude Analyses and the Relation Between Final and Intrinsic Value: **Antoine C. Dussault**

190-204 Essence et fondation: **Pablo Carnino**

205-218 Sémantique formelle et engagement ontologique: **Thibaut Giraud**

ISSN 1718-9977

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.

www.ateliers.erudit.org

La revue est financée par le Fonds de recherche sur la société et la culture (FQRSC) et administrée par le Centre de recherche en éthique (CRE)



Vous êtes libres de reproduire, distribuer et communiquer les textes de cette revue au public selon les conditions suivantes :

- Vous devez citer le nom de l'auteur et de la revue
- Vous ne pouvez pas utiliser les textes à des fins commerciales
- Vous ne pouvez pas modifier, transformer ou adapter les textes

Pour tous les détails, veuillez vous référer à l'adresse suivante :
<http://creativecommons.org/licenses/by-nc-nd/2.5/legalcode>

You are free to copy and distribute all texts of this journal under the following conditions:

- You must cite the author of the text and the name of the journal
- You may not use this work for commercial purposes
- You may not alter, transform, or build upon this work

For all details please refer to the following address:
<http://creativecommons.org/licenses/by-nc-nd/2.5/legalcode>



RÉDACTEUR 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 Mikkelson, Université McGill

COMITÉ D'EXPERTS / BOARD OF REFEREES:

Charles Blattberg, Université de Montréal

Rabah Bousbaci, 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

Béatrice Godard, Université de Montréal

Joseph Heath, Université de Toronto

Julie Lavigne, Université du Québec à Montréal

Robert Leckey, Université McGill

Bruce Maxwell, UQTR

Christian Nadeau, Université de Montréal

Pierre-Yves Néron, Université Catholique de Lille

Wayne Norman, Duke University

Ruwen Ogien, CNRS

Hervé Pourtois, Université catholique de Louvain

Bernard Reber, CNRS

Daniel M. Weinstock, Université McGill

Jurgen De Wispelaere, 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.creum.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.creum.umontreal.ca/categorie/ateliers_ethique_ethics_forum/). Papers not following these will be automatically rejected.

VERS UNE ETHIQUE CLIMATIQUE PLUS EFFICACE : MOTIVATIONS ET INCITATIONS¹

MICHEL BOURBAN

DOCTORANT EN PHILOSOPHIE, UNIVERSITÉ DE LAUSANNE ET UNIVERSITÉ DE PARIS-SORBONNE

RÉSUMÉ :

Cet article vise à justifier, puis à appliquer une éthique climatique centrée sur les intérêts des acteurs économiques. Après avoir expliqué pourquoi le changement climatique pose un problème important de motivation, je montre pour quelles raisons les incitations peuvent au moins partiellement y remédier. Je développe ensuite deux possibilités d'institutionnalisation de l'éthique des incitations. La première consiste en une taxe internationale augmentant progressivement le coût des émissions de dioxyde de carbone, un dispositif auquel il convient d'ajouter des subsides pour la recherche, le développement et le déploiement des énergies renouvelables. La seconde consiste en un marché global du carbone qui vise également à décourager l'utilisation des combustibles fossiles et à encourager l'utilisation de sources alternatives d'énergie. L'objectif est de montrer qu'une éthique climatique prenant en compte le problème de la motivation est plus efficace qu'une position qui se limite aux devoirs moraux incombant aux consommateurs et producteurs, soit de réduire leurs émissions.

ABSTRACT:

This paper tries to justify and apply a climate ethics based on the interests of the consumers and producers. First, I explain why climate change creates an important problem of motivation; then, I show how incentives can at least partially solve it; finally, I develop two different ways to institutionalize an ethics of incentives. The first market mechanism I propose is an international carbon tax increasing gradually the cost of carbon dioxide emissions, a climate policy that should be linked to subsidies for the research, development and deployment of renewable energies. The second mechanism is a cap-and-trade system, whose aim is also to disincentivize the use of fossil fuels and to incentivize the use of other sources of energy. The overall objective is to show that a climate ethics addressing the problem of motivation is more efficient than an approach dealing only with the moral duties consumers and producers have to mitigate their emissions.

INTRODUCTION

Afin de parvenir à relever le défi global et intergénérationnel posé par le changement climatique, nous devons trouver un équilibre entre quatre objectifs principaux. Du plus au moins important, ces objectifs sont : la justice ou l'équité; l'efficacité environnementale ou la maximisation des effets des politiques climatiques de réduction, d'adaptation et de compensation; le réalisme politique ou la faisabilité institutionnelle; enfin l'efficience économique ou la minimisation des coûts engendrés par les politiques climatiques. Ces éléments pouvant entrer en contradiction les uns avec les autres, leur conciliation représente une tâche complexe. Il existe même parfois des tensions au sein d'un seul objectif. Par exemple, une politique réduisant dès aujourd'hui drastiquement les émissions globales de gaz à effet de serre au nom du bien-être des générations futures peut nuire aux membres les plus défavorisés des générations présentes en augmentant radicalement les coûts de l'énergie. De même, une politique internationale attribuant trop de droits d'émission aux pays émergents peut grandement nuire aux générations futures. La justice climatique intergénérationnelle n'est pas toujours compatible avec la justice climatique globale, et vice-versa, ce qui rend le problème encore plus difficile à cerner et à résoudre.

Dans une contribution récente aux *Ateliers de l'éthique/The Ethics Forum*, Andrew Light invite les philosophes à relever les défis posés par le changement climatique de manière plus efficace que ne le font les éthiques environnementales dominantes². Cet article cherche à répondre à cette invitation en développant une éthique climatique dont les recommandations politiques sont à la fois équitables et efficaces sur le plan environnemental. Autrement dit, mon objectif est de concilier la justice et la maximisation des effets des politiques climatiques. Si je ne traite que très brièvement du réalisme politique et de l'efficience économique, ce n'est pas parce que je sous-estime leur importance; c'est tout simplement parce que la conciliation des deux premiers objectifs me semble déjà représenter une tâche suffisamment complexe.

D'emblée, je précise que je ne présente qu'une manière parmi d'autres d'accomplir cette tâche, et que les modèles d'institutions que je développe ne pourraient à eux seuls réduire les émissions globales de gaz à effet de serre. Pour y parvenir, de nombreux autres dispositifs politiques, économiques et sociaux seraient nécessaires. Il ne s'agit donc que d'une modeste contribution à la résolution du problème, mais qui me semble possiblement utile pour nous aider à sortir de l'état actuel de paralysie politique.

La première partie vise à formuler le problème de motivation posé par le changement climatique, puis à proposer une éthique permettant de le résoudre au moins partiellement : ce sera la partie la plus théorique, qui cherche à justifier l'utilisation d'incitations³ dans le cadre des politiques climatiques. La deuxième partie propose d'appliquer l'éthique des incitations sous forme d'une taxe carbone internationale couplée à un système de recherche, développement et déploiement d'énergies renouvelables. La dernière partie envisage une manière

alternative de mettre en œuvre une politique climatique incitative, en remplaçant la taxe carbone par un marché global du carbone.

ÉTHIQUE CLIMATIQUE ET ÉTHIQUE DES INCITATIONS

Philosophie et économie du changement climatique

Cela fait une vingtaine d'années que les philosophes ont commencé à publier des recherches sur le changement climatique, à peu près depuis que la science a confirmé l'existence d'une augmentation globale des températures, l'origine majoritairement anthropique de ce phénomène, et ses conséquences néfastes sur l'homme et son environnement. Si l'approche morale était et demeure dominante en éthique climatique, depuis quelques années certains auteurs cherchent plutôt à proposer des recommandations institutionnelles s'adressant directement aux preneurs de décision⁴. Au vu de l'inefficacité d'une position exclusivement morale et du peu de temps qu'il nous reste pour éviter l'émergence d'un changement climatique catastrophique, l'articulation de la phase de justification des devoirs de la justice climatique à une phase d'application de ces devoirs sous forme de politiques concrètes semble aujourd'hui effectivement incontournable⁵.

Pour donner une idée de l'état des recherches en éthique climatique, prenons pour point de départ trois articles classiques rassemblés dans le collectif *Climate Ethics: Essential Readings*⁶. En 1992, dans l'un des premiers articles philosophiques parus sur le sujet, Dale Jamieson adopte le point de vue de l'éthique des vertus pour rendre attentif au fait que si la science nous permet de comprendre toujours mieux le changement climatique, elle ne montre pas assez que notre système de valeurs est le principal responsable du problème. Selon lui, le changement climatique concerne avant tout « la façon dont nous devrions vivre et dont les humains devraient se comporter avec leurs semblables et le reste de la nature. Ce sont là des problèmes éthiques et politiques autant que des problèmes scientifiques. » Il en conclut que pour trouver une solution, il faudrait se focaliser « davantage sur le caractère moral que sur le calcul de conséquences probables⁷. »

Un an plus tard, Henry Shue aborde le problème du point de vue de la justice distributive et se demande quel critère devrait guider l'allocation des droits d'émission. À partir de la distinction entre émissions de subsistance (liées à la satisfaction de nos besoins fondamentaux, comme se nourrir, se vêtir et se loger) et émissions de luxe (liées à la satisfaction de nos besoins frivoles, comme conduire des voitures puissantes), il écrit qu'« il n'est pas équitable de demander à certains d'abandonner leurs produits de nécessité pour que d'autres puissent conserver leurs produits de luxe⁸. » Il en conclut que les droits d'émission des pays pauvres doivent être inaliénables – ils ne devraient pas pouvoir être échangés sur un marché – pour s'assurer que les riches ne puissent pas s'approprier les moyens nécessaires pour subvenir à leurs besoins fondamentaux.

En 2004, Stephen Gardiner publie un article qui commence par souligner l'interdisciplinarité du sujet, tout en relevant que « [l]e changement climatique est

un problème fondamentalement éthique. » Après avoir analysé les problèmes philosophiques soulevés par les sciences climatiques, les sciences économiques et les négociations internationales, et après s'être penché sur des questions comme le principe de précaution, la responsabilité pour les émissions passées et l'allocation des émissions futures, il conclut que « [n]ous avons un besoin urgent de clarté philosophique⁹. »

Où en sommes-nous aujourd'hui? Si les recherches scientifiques et philosophiques ont donné lieu à des publications toujours plus nombreuses, plus précises et plus inquiétantes, l'action politique fait toujours largement défaut. Tandis que le GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat)¹⁰ publie son cinquième rapport et que l'éthique climatique devient un champ de recherche de plus en plus influent, les principaux acteurs politiques et économiques démontrent une étonnante paralysie. Il existe un décalage grandissant entre, d'un côté, les connaissances scientifiques et les réflexions sur nos devoirs de justice climatique, et de l'autre, les initiatives politiques, économiques et sociales prises pour lutter contre le changement climatique.

De fait, si, à ses débuts, l'éthique climatique s'est en large partie constituée en réaction aux analyses économiques du changement climatique, aujourd'hui des rapports plus cordiaux sont possibles et souhaitables entre ces deux disciplines.

Si les articles de Shue et de Gardiner mentionnés ci-dessus critiquent chacun à leur manière, pour de bonnes raisons, certaines méthodes et certains résultats des recherches économiques, c'est l'article de Jamieson qui est le plus extrémiste. Selon cet auteur, les solutions au changement climatique ne sont en aucun cas à chercher du côté des institutions politico-économiques, étant donné qu'en la matière, « l'analyse économique conventionnelle ne sert pratiquement à rien ». En effet, « les outils d'évaluation économique ne sont pas adaptés à la tâche » et sont ainsi « condamnés à l'échec¹¹ ».

La première génération d'économistes du changement climatique, menée par Ted Nordhaus et Bjorn Lomborg, avait certes fâcheusement tendance à préconiser des politiques climatiques très faibles en raison de leurs coûts élevés par rapport à leurs bénéfices moindres (car diffus à travers l'espace et le temps). Heureusement, tous ne considèrent pas qu'il soit acceptable de sacrifier l'équité et l'efficacité environnementale sur l'autel de l'efficacité économique : la nouvelle génération d'économistes du changement climatique, menée par Nicholas Stern et Cameron Hepburn, défend des politiques bien plus substantielles, notamment grâce à l'intégration dans leurs calculs de connaissances scientifiques plus récentes ainsi qu'à la prise de conscience des fondements normatifs de leurs analyses, notamment dans les débats portant sur le taux d'actualisation social¹².

Même si ce tournant dans la recherche économique n'a pas encore marqué beaucoup de philosophes, il n'est pas pour autant passé totalement inaperçu. Simon Caney, un des philosophes du changement climatique les plus influents, a récem-

ment publié un article avec Cameron Hepburn sur le marché du carbone¹³. De même, John Broome, un économiste philosophe conciliant ces deux disciplines de manière très convaincante, vient de publier un ouvrage en éthique climatique¹⁴. Mon propos s'inscrit ici dans cet esprit de collaboration entre ces deux disciplines. Il serait naïf de la part d'un philosophe politique de vouloir développer dans le détail, à lui seul, une conception d'institutions économique-politiques, cette tâche dépassant largement son champ de compétence. Il doit plutôt rechercher l'interdisciplinarité et se contenter d'esquisser des réformes institutionnelles à partir d'un certain point de vue normatif en vue de promouvoir la justice ou de réduire les injustices.

Une structure incitative tragique

Afin de parvenir à justifier une éthique climatique des incitations, il importe tout d'abord de préciser en quoi la question des incitations est soulevée par la nature même du changement climatique. Ce problème est à la fois global et intergénérationnel. D'un côté, le lieu d'émission des gaz à effet de serre est indifférent du point de vue de leur contribution au réchauffement global : peu importe la personne, l'industrie ou le pays émettant du dioxyde de carbone ou du méthane, cette activité aura des répercussions au plan global. Cela revient à dire que tout effort de réduction d'un côté du globe peut être rendu inefficace par une augmentation égale ou supérieure d'émissions en un autre point du globe. C'est une raison majeure pour laquelle les individus, industries et États sont peu enclins à réduire leurs émissions : s'ils diminuent leur niveau de bien-être pour y parvenir, rien ne garantit que ce sacrifice ne sera pas rendu inutile par une croissance d'émissions ailleurs dans le monde. Par exemple, la province canadienne de l'Ontario utilisait encore en 2011 une demi-douzaine de centrales thermiques alimentées au charbon, qui étaient remplaçables par des infrastructures émettant beaucoup moins de gaz à effet de serre¹⁵. En même temps, la Chine construisait – et continue à construire – en moyenne deux grandes centrales au charbon *par semaine*¹⁶. Le gouvernement ontarien ayant décidé depuis lors de fermer ses centrales, ses efforts ont ainsi été rapidement annulés par l'ouverture de nouvelles centrales en Chine. Si la décision prise par les Ontariens était bien évidemment la bonne, il n'en reste pas moins que la prolifération des centrales chinoises représentait une incitation importante à ne rien faire.

Le fait que la plupart des États n'aient pas encore réduit leurs émissions est ainsi en grande partie explicable par une caractéristique centrale du changement climatique : il s'agit d'un problème d'action collective à l'échelle globale. Chaque gouvernement est incité à laisser les autres pays réduire leurs émissions pour bénéficier d'un climat plus stable sans avoir besoin d'en payer les coûts. Si l'on ajoute à cela le fait que nos systèmes sociaux, économiques et politiques – avant tout nos États-Nations souverains – n'ont pas été conçus pour résoudre de tels problèmes, l'inertie politique actuelle devient moins étonnante.

D'un autre côté, les gaz à effet de serre se dispersent également dans le temps. La durée de vie du dioxyde de carbone, le principal gaz à effet de serre anthropique, est de plusieurs siècles (entre deux et trois), avec environ 25 % de l'en-

semble des particules émises par l'homme restant dans l'atmosphère pour plus de mille ans¹⁷. Ces données font du changement climatique un phénomène sérieusement différé: ses conséquences actuelles ne reflètent pas encore les conséquences totales des gaz à effet de serre émis jusqu'à ce jour. Même si les générations actuelles parvenaient à cesser leurs émissions demain, les températures continueraient à augmenter pendant plusieurs décennies, contribuant à un réchauffement qui aurait des impacts pendant plusieurs siècles, ce qui représente un motif supplémentaire pour ne pas agir dès aujourd'hui.

Ces données soulèvent un grave problème de motivation: tandis que les générations existantes jouissent de la majorité des avantages des émissions, notamment sous forme d'énergie abondante et bon marché, les inconvénients de nos activités émettrices seront principalement ressentis par les générations futures, sous forme d'effets sévères, voire catastrophiques, du réchauffement climatique. À partir du modèle de la tragédie des communs que Stephen Gardiner reprend à Garrett Hardin, le problème peut être formulé comme une tragédie climatique intergénérationnelle¹⁸:

Collectivement, *pratiquement* toutes les générations préfèrent le résultat produit par une restriction des émissions de gaz à effet de serre par rapport au résultat produit par la surexploitation de l'atmosphère.

Individuellement, chaque génération préfère ne pas restreindre ses émissions, et prend la décision de ne pas le faire, quel que soit le choix des autres générations.

Ce qui rend le changement climatique particulièrement tragique, c'est que la plupart des membres des générations présentes vivant dans les pays développés et de plus en plus de personnes vivant dans les pays émergents (les « nouveaux consommateurs ») jouissent des bénéfices de leurs émissions et transmettent les coûts de leur mode de vie aux pauvres du monde ainsi qu'aux personnes n'existant pas encore, contribuant ainsi à empirer le phénomène. S'ils réduisaient leurs émissions, le schéma se renverserait: ils devraient supporter les coûts de leurs réductions, dont les bénéfices reviendraient aux pauvres et aux générations futures. Cette situation tragique sur le plan global et intergénérationnel incite fortement les producteurs et consommateurs polluant le plus à maintenir le statu quo.

Un autre élément d'incitation à l'augmentation des émissions globales se trouve dans certaines politiques de subsides existantes, qui viennent renforcer les incitations structurelles au changement climatique. Dans les pays développés et émergents, une bonne partie des subsides gouvernementaux dans le secteur énergétique sert à financer les combustibles fossiles, ce qui réduit leur prix et les rend beaucoup plus compétitifs que d'autres sources d'énergie ne bénéficiant pas d'un tel support. Par exemple, en 2009, les subsides pour les énergies fossiles représentaient 312 milliards de dollars à l'échelle globale, montant qui s'est élevé à 409 milliards en 2010, et qui atteindra les 660 milliards en 2020 si rien ne change¹⁹.

Loin de tenir leurs promesses solennelles de réduction d'émissions, la plupart des pays développés contribuent toujours plus au problème. En grande partie à cause de la pression de multiples groupes d'intérêts, nos gouvernements ne poursuivent pas les politiques qui restreindraient notre usage des combustibles fossiles et permettraient une transition vers une économie postcarbone. Au contraire, ils soutiennent des politiques visant l'épuisement total du pétrole, du charbon et du gaz, par tous les moyens possibles, quels que soient les dommages environnementaux. Des milliers de lobbies énergétiques parviennent ainsi à augmenter constamment les subsides pour les énergies fossiles, en réduisant chaque année la part de budget investie pour les énergies alternatives²⁰.

Justification de l'utilisation des incitations

Pour parvenir à renverser cette structure incitative tragique, il convient d'éliminer progressivement les politiques qui la renforcent et de mettre en œuvre des incitations allant dans le sens inverse. Avant toute chose, il faut justifier l'utilisation des incitations, car de nombreux philosophes ont tendance à rejeter cet instrument²¹.

Comment définir une incitation? L'alternative classique entre les différentes manières d'exercer le pouvoir (politique) oppose la force à la persuasion : contraindre les individus à accomplir ce que nous voulons, ou les convaincre de le faire volontairement. Les incitations sont une forme intermédiaire de pouvoir, que l'on peut définir comme un coût ou un bénéfice représentant un découragement ou un encouragement à l'action, qui vise à modifier de manière délibérée le comportement d'un agent ou d'un groupe d'agents.

Cet outil soulève des problèmes éthiques, étant donné qu'il peut être utilisé de manière illégitime²². Pour illustrer un type immoral d'incitation, en partant du rôle de la croissance démographique dans l'augmentation des émissions globales, certaines politiques pourraient chercher à réguler le nombre de naissances en éliminant des aides de l'État ou en créant de nouvelles taxes. L'État pourrait par exemple instaurer une taxe pour le mariage et pour chaque enfant, éliminer les congés de maternité ou les allocations familiales, accorder des pensions aux femmes d'un certain âge n'ayant pas d'enfants, ou encore éliminer les aides publiques pour l'assurance-maladie, l'éducation et le logement pour toute famille ayant plus d'un enfant²³. Des droits obtenus au prix de décennies, voire de siècles de luttes pour plus de justice sociale se verraient ainsi remis en question. L'aspect le plus injuste de ce type de politique est qu'il toucherait avant tout les membres les plus défavorisés de la population, qui dépendent des aides financières de l'État pour parvenir à mener une existence satisfaisante. Les effets d'une telle politique seraient désastreux : des réductions d'émissions pourraient certes être obtenues, mais cette pratique créerait bien plus de problèmes qu'elle n'en résoudrait. L'objectif est de trouver un équilibre entre l'équité et l'efficacité environnementale, non de sacrifier la première au profit de la seconde.

Les incitations ont comme avantage de pousser les individus à l'action là où le devoir moral et la vertu échouent à les faire changer d'attitude : elles permettent

ainsi de s'attaquer au problème de la motivation, qui est, comme nous venons de le voir, très saillant dans le cas du changement climatique. Leur désavantage est qu'elles ne convainquent pas les agents y étant soumis : la persuasion est une forme de pouvoir plus légitime, notamment dans les démocraties, où l'on justifie le plus souvent les décisions politiques par le consentement direct ou indirect de la majorité. Les incitations sont une forme de manipulation des citoyens, raison pour laquelle il faut les utiliser avec parcimonie. Néanmoins, dans le cas du changement climatique, elles sont le plus souvent légitimes, et je voudrais montrer pourquoi il en est ainsi avant d'expliquer comment nous pourrions les mettre en œuvre.

Comme nous venons de le voir, il existe déjà une structure d'incitation : le changement climatique est une tragédie globale et intergénérationnelle. L'unique choix s'offrant à nous est de conserver cette structure tragique comme telle ou de chercher à la modifier. La question ici n'est pas de savoir s'il faut inciter ou non, mais plutôt de savoir comment le faire au mieux. L'alternative se trouve entre continuer à stimuler la tragédie climatique globale et intergénérationnelle en incitant les consommateurs et producteurs à augmenter leurs émissions par des subsides sur les énergies fossiles, ou chercher à éviter l'émergence d'un changement climatique catastrophique en finançant des sources alternatives d'énergie tout en augmentant le prix des combustibles fossiles.

Une autre raison pour laquelle l'utilisation des incitations est justifiée est que la persuasion est insuffisante : elle ne représente pas une forme efficace de politique climatique. C'est la politique favorite de Dale Jamieson, pour qui « la seule solution à ce problème²⁴ » consiste en l'éducation à une vie bonne aboutissant à l'action politique d'un mouvement de citoyens vertueux. Le changement climatique étant doté d'une structure incitative tragique, l'éducation ou l'argumentation morale ne peuvent pas suffire à convaincre les individus à changer de comportement : malgré plus de vingt années de recherches, publications, conférences et enseignements en éthique climatique, c'est toujours l'inaction qui domine. Autrement dit, outre les raisons morales pour combattre le changement climatique, il faut également tenir compte des raisons amORALES, relevant de l'intérêt des producteurs et consommateurs.

Prenons un exemple de mesure volontaire de réduction d'émissions développé par Andrew Szasz afin d'explicitier pourquoi la persuasion est insuffisante : la consommation verte, qui consiste à choisir des biens et services à faible densité carbone²⁵. Pour que cette pratique devienne répandue, plusieurs conditions doivent être réunies. Les consommateurs doivent savoir que le changement climatique existe, qu'il est causé par les émissions anthropiques de gaz à effet de serre et qu'il est urgent d'y remédier, ce qui est loin d'être le cas, notamment aux États-Unis, où les campagnes de désinformation ont eu et continuent à avoir des effets désastreux. Il faut également que les biens et services à faible densité carbone comme les voitures hybrides, l'électricité provenant d'énergies renouvelables ou encore un réseau dense et efficace de transports publics soient non seulement disponibles, mais le soient à des prix accessibles, ce qui n'est pas

assez souvent le cas. Presque tous les aspects de notre vie quotidienne étant directement ou indirectement associés à des émissions de gaz à effet de serre, ces biens et services doivent en outre être disponibles et accessibles pour une quantité très importante d'activités quotidiennes. Une dernière condition a trait au problème de la motivation : d'autres motifs que la conscience environnementale entrent en jeu dans les choix du consommateur, comme certaines valeurs culturelles, la rivalité de niveau social avec ses voisins et la pression des proches.

Szasz conclut que la plupart de ces conditions ne sont pas réunies aujourd'hui, et que sans des taxes et des subsides pour pousser les acteurs économiques à changer leurs habitudes, la consommation verte ne sera jamais répandue. Il ne s'agit pas de dire que la modification des comportements n'est pas une tâche importante; il s'agit simplement de se demander comment parvenir à la réaliser au mieux. Attendre des individus qu'ils fassent ce choix volontairement en les y convaincant ou en les y éduquant prend trop de temps. Le changement climatique ne pourra être maintenu à un niveau tolérable pour la plupart des êtres humains que si nous parvenons à réduire les émissions globales dans les années, au maximum dans les quelques décennies à venir, pour éviter l'émergence, dans un futur plus ou moins proche, de phénomènes catastrophiques déclenchés par un réchauffement climatique abrupt. Light nous met en garde : « pour échapper à certains des pires scénarios futurs de changement climatique, les climatologues estiment que nous devons atteindre le niveau maximal d'émissions globales cette décennie [...]. Si les philosophes souhaitent participer aux efforts pour empêcher un réchauffement catastrophique, ils doivent mettre de côté les projets à long terme décrits par Arne Naess, et chercher plutôt à influencer les institutions qui sont développées ici et maintenant pour travailler à la création d'une solution²⁶. »

Un dernier argument en faveur de la mobilisation des incitations est que si ce pouvoir n'est certes pas strictement persuasif, il n'est pas pour autant totalement contraignant : en modifiant le prix des diverses sources d'énergie à la disposition des individus, nous ne les forçons pas à choisir une source plutôt qu'une autre; nous les poussons simplement dans une direction plutôt qu'une autre, tout en leur laissant en dernière instance le choix de réduire leurs émissions ou non. Tant les marchés du carbone que les taxes permettent aux entités régulées (que ce soient des pays, des entreprises ou des individus) de réduire librement leurs émissions en fonction de la méthode qui leur semble la plus adéquate. La nouvelle situation se matérialisant grâce à l'utilisation des incitations est qu'à chaque fois que ces entités décident d'émettre davantage, elles doivent dépenser plus pour couvrir les coûts sociaux et environnementaux découlant de leurs choix. Dans le jargon des économistes, les externalités sont internalisées, l'ensemble des coûts des combustibles fossiles est intégré à leur prix.

Il ne s'agit certes plus ici d'une liberté au sens le plus exigeant du terme, à savoir l'autonomie du citoyen qui choisit sans influence de l'État les normes auxquelles il soumet son action; mais si nous en restons à cette conception de la liberté, d'importantes injustices climatiques continueront à émerger, raison pour laquelle il faut que la marge de manœuvre des producteurs et consommateurs soit réduite,

dans le secteur de l'énergie, à une liberté de choisir entre les sources plus ou moins polluantes, tout en les incitant à choisir ces dernières. Le citoyen n'est certes plus considéré ici comme un agent moral pleinement responsable, mais l'usage d'un certain degré de coercition par l'État afin de protéger les personnes des conséquences néfastes des actions d'autrui fait partie des instruments politiques permettant de garantir la justice. Si les agents moraux ne respectent pas leurs devoirs de justice, dans ce cas « il est légitime de réduire leur autonomie en les forçant à respecter les exigences de la justice²⁷. » D'où l'importance de mettre en place un dispositif incitatif visant non à remplacer, mais à compléter les mesures volontaires ou autonomes de réduction d'émissions prises par les individus comme le choix d'un mode de vie moins dense en énergies fossiles, dont la consommation verte n'est qu'une facette²⁸.

VERS UNE TAXE CARBONE INTERNATIONALE

Si l'on veut parvenir à réduire les émissions de telle manière à éviter un changement climatique abrupt, « [l']analyse économique montre que seule la mise en place, par les autorités, d'un signal de prix adéquat permet d'engendrer les incitations qui guident les choix des différents agents économiques – producteurs et consommateurs – vers les options compatibles avec un tel sentier de réduction²⁹. » Depuis les études de l'économiste Arthur Pigou, nous savons que les externalités causent des défaillances de marché. Or, les émissions de gaz à effet de serre sont des externalités qui représentent la plus grande défaillance de marché jamais atteinte auparavant³⁰. Pour y remédier, il faut implanter des instruments permettant de renverser la structure existante des incitations : il s'agit de mettre sur pied des institutions permettant de rendre les émissions plus coûteuses et les énergies alternatives plus compétitives. Dans les termes du *Rapport mondial sur le développement humain 2007-2008*, « [l]e point de départ consiste à affecter un prix aux émissions de carbone. La modification des structures d'incitation est une condition vitale pour accélérer la transition vers une croissance à faibles émissions de carbone³¹. »

Le recours à l'intervention publique est incontournable, que ce soit sous forme de réglementation (fixer des normes d'émission), de taxation pigouvienne (intégrer dans la valeur marchande le coût social et environnemental des externalités) ou de marché du carbone (distribuer des droits d'émission échangeables au sein d'un marché dans lequel s'établit un prix augmentant le coût des activités émettrices). Dans ce qui suit, je développerai les deux dernières options, puisqu'elles représentent deux manières d'appliquer l'éthique climatique des incitations. La réglementation est également une solution importante, mais elle ne recourt pas aux incitations, raison pour laquelle je ne la prends pas en compte ici.

Réduction des subsides pour les énergies fossiles et création d'une taxe carbone

L'objectif d'une réforme institutionnelle passant par les systèmes de taxation domestique des différents pays membres de la CCNUCC (Convention-cadre des Nations Unies sur les changements climatiques) est de remplacer la logique, lar-

gement insuffisante et inefficace, de cibles absolues de réduction d'émissions utilisée actuellement dans les négociations internationales par une logique d'incitation par le prix. L'idée est de remplacer les promesses volontaires et non contraignantes des gouvernements à un pourcentage de réduction de leurs émissions nationales par le prélèvement d'un prix sur la tonne de CO₂. Aucun pays ne subirait de réduction de sa compétitivité économique, les autres États mettant en œuvre le même type de politique sur leur territoire en fonction de leur responsabilité causale pour leur problème et de leur capacité d'y remédier. L'efficacité économique serait ainsi prise en compte, les problèmes de compétitivité internationale étant réduits en vertu du fait que chaque pays mettrait en œuvre une politique similaire. L'équité serait également garantie, la taxation tenant compte non seulement du niveau des émissions, mais également de la répartition des revenus dans les différents pays³². S'il revient aux économistes et aux scientifiques d'établir le montant exact de cette taxe, le philosophe peut contribuer au débat en justifiant une certaine utilisation des incitations et en s'assurant que la mise en œuvre d'un « signal-prix » mondial de carbone respecte les normes d'équité. Autrement dit, le rôle du philosophe est de compléter, et si nécessaire, corriger l'approche souvent descriptive de l'économiste par une approche normative permettant de juger si la politique proposée est éthiquement acceptable. Ce type d'exercice est d'autant plus pertinent que de plus en plus d'économistes du changement climatique s'efforcent d'intégrer des exigences normatives dans leurs raisonnements, ce qui ouvre la porte à ce type de collaboration interdisciplinaire.

Le premier pas consiste à réduire rapidement les subsides pour les combustibles fossiles. Cette stratégie est nécessaire, puisque sans elle une taxe carbone n'aurait aucun sens : « faire l'un sans l'autre serait intrinsèquement contradictoire³³. » Une raison supplémentaire pour réduire ces subsides est qu'ils sont très inéquitables, bénéficiant avant tout aux plus riches. En 2009, environ 80 % des avantages procurés par les subventions aux produits pétroliers ont profité aux ménages les plus riches, tandis qu'en 2010, seulement 8 % de ces subventions ont profité au quintile de la population le plus pauvre³⁴. De telles subventions ont en outre pour effets d'accélérer la surconsommation et l'épuisement des ressources, de fausser les signaux du marché, de faire peser sur les gouvernements un lourd fardeau financier, de décourager l'investissement dans les infrastructures énergétiques alternatives et d'entraver l'adoption de pratiques et de procédés plus éco-énergétiques. Il est économiquement insensé de gaspiller des ressources financières publiques rares et précieuses pour subventionner ce qui représente l'industrie la plus profitable de l'histoire du capitalisme.

Cette mesure permettrait certes de réduire quelque peu l'utilisation des énergies fossiles par les consommateurs et producteurs, mais puisque nous sommes ici en présence d'un des secteurs de l'économie les plus prospères qui ait jamais été, les investissements privés remplaceraient rapidement les investissements publics si aucune autre mesure n'était prise. Pour cette raison, nos économies n'ont rien à craindre d'une annulation des subsides de l'État; mais en même temps, les émissions globales continueraient à croître.

Afin d'y remédier, le second pas consiste à mettre en œuvre une taxe carbone internationale dont le montant augmentera graduellement pour s'assurer que le prix des combustibles fossiles croisse progressivement et que leur compétitivité sur le marché soit réduite par rapport aux sources alternatives d'énergie. La taxe ne doit pas être initialement trop haute, pour laisser le temps aux producteurs et consommateurs de s'adapter aux nouvelles énergies et ne pas créer une récession économique ; mais elle ne doit pas non plus être trop basse, sinon les incitations par le signal du prix seraient trop faibles pour orienter les individus vers d'autres sources d'énergie. Une mesure importante pour y parvenir revient à préparer dès aujourd'hui les principaux acteurs du marché à utiliser les énergies renouvelables, notamment à travers le management durable³⁵. De plus, ne pas attribuer un prix au carbone revient à subsidier de manière informelle les combustibles fossiles, les acteurs économiques pouvant continuer à créer des externalités sans devoir en payer les coûts : pour mettre fin à toutes les formes de subsides sur le pétrole, le gaz et le charbon, la création d'une taxe (ou, comme nous le verrons plus bas, d'un marché du carbone) est donc une étape nécessaire³⁶.

Subsides pour les énergies renouvelables et segmentation des marchés

Ce dispositif ne fonctionnera que si des sources alternatives d'énergies sont disponibles et accessibles pour les individus et entreprises qui décident de changer leurs habitudes à cause de l'augmentation du prix du pétrole, du charbon et du gaz. Ici, l'État peut jouer un rôle déterminant en investissant dès aujourd'hui massivement dans la recherche, le développement et le déploiement d'énergies renouvelables comme le solaire, l'hydraulique et l'éolien. Cet investissement considérable peut être principalement financé par les subsides qui ne sont plus investis dans les combustibles fossiles et par les fonds soulevés par la taxe carbone. Une autre source envisageable serait la réduction des investissements de l'État dans d'autres secteurs, par exemple dans l'armée, le changement climatique devenant de plus en plus un problème de sécurité nationale³⁷.

Une fois ces technologies développées, les pays nantis ne devront pas se contenter de les déployer sur leur propre sol ; ils devront également s'assurer que les pays émergents ainsi que les pays pauvres puissent y avoir accès par des transferts de technologie et des politiques de brevet plus souples que celles qui existent. Si les brevets sont un instrument permettant d'inciter les chercheurs à développer de nouveaux projets, un système trop exigeant de monopole sur les inventions utiles à l'humanité empêche les pauvres du monde de pouvoir en profiter, comme cela s'observe notamment dans le cas des brevets sur les médicaments sous l'Organisation mondiale du commerce³⁸. Pour réduire les effets pervers du monopole temporaire rendu possible par les brevets sur les nouvelles technologies, la méthode de la segmentation des marchés peut s'avérer très utile : les titulaires de brevets sur les technologies vertes peuvent accepter de vendre leurs produits à un prix réduit aux pays en développement, tout en maintenant des prix plus élevés pour les pays développés³⁹. Il ne s'agit donc pas d'imposer une politique de développement aux pays en développement, mais de leur garan-

tir l'accès à des sources alternatives d'énergie lorsqu'elles sont moins coûteuses que les combustibles fossiles, puisque tant que des pays comme la Chine et l'Inde continueront à produire la majeure partie de leur électricité en utilisant du charbon, aucune solution au changement climatique ne sera envisageable. Il est également possible que les technologies de demain voient le jour dans un pays émergent : dans ce cas, les transferts de technologie vers les pays pauvres et la segmentation des marchés entre les pays développés et les autres pays émergents doivent être maintenus, pour les mêmes raisons.

Politiques compensatrices et ajustements aux frontières

Afin d'obtenir une politique équitable, il convient également d'anticiper ses effets sur les plus démunis. L'équité peut ici être entendue de deux manières. Elle peut soit chercher à éviter que les politiques climatiques aient des effets néfastes sur les personnes les plus défavorisées selon un principe de non-nuisance, soit viser en outre l'augmentation du bien-être de ces personnes selon un principe de redistribution des ressources permettant de les enrichir. Si la conception maximaliste qui vise à redistribuer les richesses des plus aisés vers les plus démunis est un objectif préférable du seul point de vue de la justice, la conception minimaliste qui se limite à une protection des victimes du changement climatique ainsi que des populations et individus pouvant souffrir des dispositifs de réduction d'émission est plus réaliste et représente un meilleur compromis entre l'équité et l'efficacité. Empêcher des prises de décision politique en considérant la quête d'une solution idéale plus importante ne rend pas seulement les recommandations philosophiques inefficaces et irréalistes, mais également, à la longue, inéquitables, les victimes du changement climatique qui auraient pu profiter de solutions moins ambitieuses que celles valant uniquement dans un monde idéal se retrouvant au final sans protection. C'est pour cette raison que la justice et l'efficacité peuvent être, sous certaines conditions, des objectifs conciliables : sans la seconde, la première peut très bien ne jamais être mise en œuvre, une théorie prescrivant des cours d'action trop exigeants ou sacrificiels n'ayant que très peu de chance d'inspirer les acteurs économiques et politiques; sans la première, la seconde risque d'être rejetée, trop de preneurs de décision politique considérant qu'une solution peu équitable, voire injuste, ne peut être acceptée, ce qui entraîne dans les deux cas le maintien et la radicalisation des injustices climatiques.

Pour parvenir à remplir les exigences de la conception minimaliste, il faut ajouter à notre dispositif des mécanismes de compensation. Une taxe carbone a pour effet pervers d'appauvrir les ménages qui ont le moins de revenus et de richesses : l'augmentation du coût des énergies fossiles touche avant tout les personnes et familles à faible revenu, pour qui une part plus importante du budget est consacrée aux dépenses énergétiques. Aussi, la taxe peut facilement devenir régressive si des mesures compensatrices ne sont pas prises. Par exemple, le gouvernement pourrait mettre en place des subsides visant à améliorer l'efficacité énergétique dans les ménages défavorisés : la mise en place d'infrastructures à faible consommation d'énergie leur coûterait moins que l'adoption de leur analogue à faible efficacité énergétique, et de larges économies sur leur bud-

get énergétique seraient possibles en vertu de leur niveau très bas de consommation. De même, des subsides pourraient être investis dans le domaine des transports publics en vue d'agrandir leur réseau, d'améliorer leur efficacité et de réduire leur prix, l'augmentation du coût des hydrocarbures ayant un impact important sur les classes populaires et moyennes au plan des transports privés⁴⁰.

Un autre mécanisme possible consisterait à redistribuer une partie des recettes collectées par la réduction des subsides pour les énergies fossiles et par la taxe carbone sous forme d'un dividende compensant l'augmentation du coût de la vie. Vu que les consommateurs les plus fervents des énergies fossiles sont les plus riches, l'impôt aurait ainsi un effet compensateur en permettant une redistribution des richesses des plus aisés vers les plus pauvres, ce qui garantirait que la taxe ne soit pas régressive. Ici aussi, les spécificités de la mise en œuvre sont à préciser par les économistes, mais l'essentiel est de s'assurer que cette mise en œuvre soit guidée par une conception au moins minimaliste de la justice.

Un dernier aspect est celui des pays qui ne respectent pas leur engagement ou refusent de rejoindre l'accord international sur la fixation d'un signal prix carbone. Une mesure envisageable pour éviter l'émergence de ce problème de passer clandestin consiste à tarifier aux frontières les produits en provenance des pays sans prix carbone conforme à la norme internationale par un prélèvement équivalant la taxe appliquée au produit national. Si l'on essayait d'appliquer ce dispositif à toutes les importations, le commerce international pourrait progressivement se bloquer et nous nous trouverions très rapidement devant l'impossibilité de calculer la teneur en carbone de chaque bien particulier. C'est pour cela que la taxe à la frontière devrait se restreindre aux biens qui sont actuellement responsables de la plupart des émissions importées, comme l'acier, les produits chimiques, l'aluminium, le ciment et les engrais. Il ne s'agit pas là d'une mesure protectionniste; c'est au contraire le fait de ne pas appliquer une taxe carbone qui représente une distorsion du prix, puisque cela revient à subsidier des exportations polluantes.

Un ajustement aux frontières permettrait de transmettre le message suivant aux principaux émetteurs : si vous ne prélevez pas un prix carbone sur vos produits avant de les exporter, les autres pays le feront à votre place, et ce sont eux qui en tireront des revenus. Aussi, chaque pays aurait un fort intérêt commercial à se conformer à l'accord international. Il s'agit ici aussi d'une incitation puissante à la réduction des émissions de gaz à effet de serre, puisque les gouvernements préfèrent taxer eux-mêmes leurs propres biens plutôt que de les voir taxés ailleurs. Ce mécanisme permettrait également de rendre le traité climatique futur plus contraignant que le protocole de Kyoto : en l'absence d'un gouvernement mondial imposant à chacun des membres de la CCNUCC le respect de leurs engagements, le recours aux sanctions économiques représente une manière de pousser les passagers clandestins potentiels à se conformer aux exigences du traité climatique.

Dieter Helm en conclut que « la taxe carbone à la frontière correspond à une logique ascendante incitant les autres pays à prendre part à la création d'un prix carbone global⁴¹. » Au lieu de partir d'une logique descendante inefficace d'imposition de pourcentages de réduction d'émissions nationales, cette taxe suit une logique ascendante d'attribution d'un prix aux biens produits à l'extérieur, mais consommés à l'intérieur d'un pays, qui peut partir d'un nombre limité d'États, comme les membres de l'UE, puis se répandre au reste de leurs partenaires commerciaux.

VERS UN MARCHÉ DU CARBONE GLOBAL

Si le dispositif que je viens de proposer me semble remplir les deux objectifs de l'équité et de l'efficacité environnementale, on peut légitimement se demander s'il ne manque pas le troisième objectif principal en matière de politique climatique : la faisabilité institutionnelle. En partant des trois premiers objectifs que j'ai présentés dans mon introduction, on pourrait m'objecter que ma proposition ne respecte pas le troisième : n'est-il pas irréaliste de défendre une taxe carbone internationale, alors que cette institution rencontre actuellement une forte opposition dans de nombreux pays ?

Si une telle politique échoue, certes, aux plans national et international, elle peut être réalisée avec succès au plan subnational, même en l'absence de politiques climatiques domestiques ou d'un traité international climatique efficace : cette réponse locale ou régionale au changement climatique est effectivement en plein essor depuis les échecs répétés des négociations internationales⁴². Le problème est que cette solution est peu efficace : les dispositifs subnationaux n'ont jusqu'ici eu que très peu d'effets sur les émissions globales, et si nous voulons parvenir à stabiliser, puis réduire la concentration atmosphérique de gaz à effet de serre, la taxe devra également être mise en œuvre au niveau national dans les pays qui émettent le plus.

Une autre réponse consiste à se pencher sur un autre type d'exercice d'application de l'éthique des incitations. Si le but des politiques climatiques incitatives est d'augmenter le prix des énergies fossiles et de réduire celui des énergies renouvelables, alors un marché global du carbone devrait être au moins tout aussi performant, et même plus réaliste. C'est ce que je voudrais brièvement essayer de montrer avant de conclure.

Théorie des systèmes de plafonnement et d'échange

Les systèmes de plafonnement et d'échange, ou marchés du carbone, sont une institution permettant également d'attribuer un prix aux énergies fossiles, mais en utilisant un mécanisme différent de celui d'une taxe carbone. Ce dispositif établit un plafond sur les émissions régionales, nationales ou internationales de gaz à effet de serre pour déterminer la quantité de droits d'émission pouvant être distribués sur le marché. Par exemple, au plan global, le plafond représenterait le stock total de gaz à effet de serre que l'atmosphère peut absorber sans déclencher un changement climatique catastrophique. Une fois ce quota établi, les participants reçoivent une autorisation à s'engager dans les activités émettrices

soumises au mécanisme sur une période donnée : les systèmes en amont se focalisent sur un nombre limité de fournisseurs, importateurs et raffineurs d'énergie, tandis que les systèmes en aval se focalisent sur les principaux consommateurs d'énergie, que ce soient des individus, des ménages, des entreprises ou des institutions publiques. Une quantité fixe et graduellement réduite de permis d'émissions peut ensuite être mise à disposition sur le marché : ils sont soit distribués gratuitement, soit vendus aux enchères.

Dès que le système est opérationnel, les participants peuvent vendre les droits d'émission dont ils n'ont pas besoin sur le marché ou en acheter s'ils n'en ont pas assez : toute entreprise économisant de l'énergie ou changeant de source d'énergie peut ainsi vendre ses droits supplémentaires, et toute entreprise polluante plus doit payer davantage. Les acteurs économiques émettant plus de gaz à effet de serre qu'ils n'y ont droit sont soumis à des sanctions financières et légales : pour s'assurer que les participants se conforment aux règles du mécanisme, il faut pénaliser ceux qui ne les respectent pas⁴³. Ici aussi, le recours aux sanctions économiques permet de garantir l'efficacité du dispositif.

Un tel système comprend plusieurs avantages notables. Il jouit d'un degré élevé de faisabilité institutionnelle, le marché du carbone étant déjà une pratique largement répandue. Pour avoir une chance d'être rapidement adoptées, les recommandations politiques doivent partir de là où nous nous trouvons, plutôt que de là où nous voudrions être⁴⁴ : les marchés du carbone respectent cette contrainte, puisqu'ils ont non seulement été adoptés au niveau international par l'UE, mais aussi au niveau national dans de nombreux pays comme la Suisse et la Nouvelle-Zélande, ainsi que dans de nombreuses entités subnationales, comme aux États-Unis et au Canada. Cette exigence est donc mieux respectée que dans le cas de la taxe carbone. De même, le plafonnement garantit qu'un objectif précis de réduction d'émission sera atteint, alors qu'avec une taxe carbone cet objectif demeure indéterminé.

Un autre avantage à relever est que, tout comme dans le cas de la taxe carbone, les individus conservent leur liberté de choisir entre différentes sources d'énergie, tout en étant incités à choisir la moins polluante. En fixant une quantité maximale d'émissions, puis en distribuant les droits d'émission correspondant à la différence entre la quantité actuelle de gaz à effet de serre dans l'atmosphère et le plafond fixé, ce mécanisme permet de rendre les émissions plus coûteuses et d'inciter les producteurs et consommateurs à se tourner vers des sources d'énergie alternatives. Le marché du carbone diminue ainsi les chances de l'émergence d'un changement climatique catastrophique grâce à la manière dont il modifie le comportement des agents économiques.

Mais cette institution, comme tous les mécanismes de marché, pose également des problèmes éthiques non négligeables. Plusieurs philosophes s'y opposent, estimant que les droits d'émission représentent une propriété privée de l'atmosphère, alors qu'il s'agit d'un service environnemental ne pouvant être privatisé ; une marchandisation de l'atmosphère, qui est réduite à un bien de

consommation parmi d'autres; ou encore une forme d'indulgence, qui permet aux plus riches de payer pour être exemptés de leurs devoirs de justice climatique. Ces objections théoriques sont importantes, et essayer d'y répondre ici prendrait trop de place⁴⁵; je tiens néanmoins à répondre à un autre argument important, qui concerne le problème de la motivation.

Selon Andrew Dobson⁴⁶ et Clive Spash⁴⁷, les politiques fondées sur le marché diminuent la motivation intrinsèque des agents moraux à réduire leurs émissions, puisqu'ils ne perçoivent plus le respect de l'environnement comme gratifiant en soi, mais comme source de gains financiers. La création d'un marché sur lequel nous pouvons échanger nos droits d'émission promeut la protection de l'environnement pour des raisons instrumentales plutôt qu'intrinsèques. La conséquence contre-productive en est que les incitations par le prix réduisent la motivation morale des individus à lutter contre le changement climatique. Tant les taxes carbone que les systèmes de plafonnement et d'échange poussent à agir rationnellement, à la place de laisser l'agent moral faire le libre choix d'un comportement vertueux.

Pour que cette objection soit valable, certaines conditions doivent être réunies. Par exemple, la conscience environnementale doit être largement répandue, tant au niveau des individus que des entreprises et des États. Or, comme Edward Page le relève, « le problème ici est que, tout simplement, la plupart des gouvernements, des millions d'entreprises et des milliards d'individus semblent avoir à l'origine peu ou aucune conscience écologique à remplacer [...] »⁴⁸. Les droits d'émission sont en outre le plus souvent distribués aux entreprises : l'éthique étant un facteur bien moins important pour une entreprise (dont le rôle principal est de créer de la richesse, non de la vertu, même si pour le faire elle doit évidemment respecter un certain code déontologique) que pour un individu, peu de valeurs intrinsèques risquent d'être subverties.

Un autre problème posé par cette objection est que les individus, entreprises et États déjà convaincus de l'importance de la lutte contre le changement climatique ont au contraire toutes les raisons de se réjouir de la mise en place d'institutions motivant les autres à agir dans la même direction : cela réduirait le risque que leurs réductions d'émission soient rendues inefficaces par une augmentation ailleurs. Et même s'ils trouvent un intérêt financier à réduire leurs émissions, en principe, cela ne leur fera pas perdre leur souci de l'environnement, les deux n'étant pas exclusifs.

Finalement, même si dans certains cas la motivation intrinsèque se voit remplacée par une motivation instrumentale, cela n'importe presque pas si l'objectif est de réduire les émissions globales : l'important n'est pas de réduire nos émissions parce que nous y voyons une responsabilité civique ou un gain financier, mais que nous le fassions avant qu'il soit trop tard, peu importe la nature de la motivation nous y poussant. C'est ici que les limites de l'éthique des vertus et de l'éthique déontologique se font le plus sentir, et que les avantages d'une éthique des incitations sont le plus manifestes : que nous soyons des citoyens

vertueux ou des consommateurs et producteurs suivant notre propre intérêt, peu importe tant que les émissions globales sont stabilisées, puis réduites⁴⁹.

Pratique des systèmes de plafonnement et d'échange

Les problèmes pratiques posés par les systèmes de plafonnement et d'échange sont tout aussi importants que les problèmes théoriques, raison pour laquelle cette section ne prétend pas non plus faire le tour de la question. Par exemple, je laisse de côté les questions relatives à la nature du porteur de droit (les individus, les entreprises ou les États?), à l'étendue et la nature des activités économiques soumises au plafonnement (seulement les activités qui génèrent de hautes émissions ou toutes les activités émettrices; plutôt les activités liées à la production ou plutôt celles liées à la consommation?), ou encore les questions portant sur l'évaluation du quota d'émissions (quel seuil de dangerosité ne doit pas être dépassé, au-delà duquel le changement climatique risque de devenir catastrophique?). Toutes ces questions impliquent des réflexions éthiques dignes d'être approfondies, mais elles dépassent la portée de cet article. Je vais pour cette raison plutôt me concentrer sur les problèmes pratiques qui sont en relation directe avec mon objectif.

Tout comme la taxe carbone, les systèmes de plafonnement et d'échange peuvent avoir un effet régressif: lorsque nous augmentons les coûts de la pollution, nous prenons le risque de toucher avant tout les plus pauvres. Pour éviter cela, les droits d'émission devraient être vendus aux entreprises par les États, les gains ainsi réalisés servant à compenser l'augmentation des prix de l'énergie pour les plus pauvres. Dans plusieurs systèmes existants, beaucoup trop de droits d'émission sont distribués gratuitement; ils sont même parfois alloués en fonction du niveau actuel d'émission des entreprises, donnant plus de droits aux plus polluantes, et moins à celles qui émettent le moins, récompensant ainsi celles qui ont le plus contribué au problème. Ces deux défauts majeurs sont dus à l'influence des intérêts industriels sur les politiques climatiques existantes. Ils peuvent être estompés si les permis d'émissions sont vendus, ce qui pénalise les entreprises qui polluent le plus et récompense celles qui polluent le moins, tout en créant un fonds destiné à neutraliser l'effet régressif du système et à investir dans les énergies renouvelables.

Afin de rendre le marché du carbone plus équitable et plus efficace, un autre principe de distribution que celui du niveau actuel d'émission des entreprises doit également être choisi. Ici se pose la question du critère de l'allocation des droits d'émission: selon quel principe de justice doit-on distribuer les droits d'émissions? Laisant ici de côté le débat très riche de philosophie politique sur le meilleur critère du point de vue de la justice, je voudrais simplement me demander quel est le minimum requis par l'équité pour garantir que la distribution des droits d'émission ne se fasse pas aux dépens des populations et personnes les plus défavorisées. Comme dans mon analyse de la taxe carbone, je me borne donc à une conception très modeste (c'est-à-dire incomplète) de la justice.

À son niveau minimal, la justice exige que chacun ait assez pour pouvoir mener une vie décente, qui ne se situe pas sous un seuil absolu de suffisance, ce qui implique notamment que les besoins les plus fondamentaux (comme se nourrir, se vêtir et se loger) soient satisfaits. Selon Harry Frankfurt, un fameux défenseur de cette approche suffisantiste, le seuil est même plus élevé qu'une vie simplement décente, étant donné qu'elle doit également être satisfaisante⁵⁰. L'important est que la distribution des droits d'émission garantisse au moins que les besoins fondamentaux des personnes d'aujourd'hui et demain soient satisfaits.

Un bon critère pour s'assurer que cette exigence minimale de la justice soit respectée revient à distribuer les permis d'émission de telle manière que les droits fondamentaux des individus soient protégés. Prenons par exemple le droit à la santé : pour que les personnes puissent jouir d'un standard minimal de santé, une quantité importante d'électricité est nécessaire pour créer et entretenir les infrastructures comme les cliniques : pouvoir utiliser les instruments médicaux adéquats, réfrigérer les vaccins, avoir de la lumière pour s'occuper des patients la nuit, améliorer l'hygiène, pouvoir communiquer et être au courant des événements extérieurs, etc. De même, pour garantir le droit à la subsistance, une certaine quantité d'électricité est à nouveau nécessaire pour produire et cultiver la nourriture, la transporter, la stocker, la cuisiner, etc. Dans les deux cas, l'énergie est le plus souvent fournie par les combustibles fossiles, surtout dans les pays pauvres où les types alternatifs d'énergie coûtent trop cher ou n'existent tout simplement pas. Il est désirable d'aider ces pays à développer ces autres types d'énergie, mais dans l'état actuel des choses, la priorité est de s'assurer qu'ils aient au moins assez de droits d'émission pour pouvoir subvenir à leurs besoins fondamentaux : « [e]n identifiant les droits des personnes et leurs implications en termes d'émission de gaz à effet de serre, on peut en déduire la manière dont ces émissions devraient être distribuées⁵¹. »

Aussi longtemps que les personnes auront besoin des énergies fossiles pour subvenir à leurs besoins fondamentaux, la distribution des permis d'émission ne devra pas être trop coûteuse pour les personnes qui peinent déjà à survivre. Afin de garantir le niveau minimal d'émissions nécessaires à une vie décente, ceux qui sont le moins capables de payer les droits d'émission doivent les recevoir gratuitement. Sans cela, la justice climatique intergénérationnelle risque d'entrer en contradiction avec la justice climatique globale : si, au nom du bien-être des générations futures, nous ne distribuons pas gratuitement les droits d'émission aux plus pauvres, alors nous risquons de les soumettre à un mécanisme auquel elles ne peuvent tout simplement pas s'ajuster⁵².

Henry Shue fait partie des rares philosophes à avoir constaté l'attrait des incitations et des mécanismes de marché. Tout comme les autres auteurs n'étant pas de formation philosophique cités ci-dessus (Éloi Laurent, Dieter Helm, les rédacteurs du rapport du PNUD, etc.), il part du constat selon lequel « toute méthode viable de réduction d'utilisation des combustibles fossiles dépend de l'élévation de leurs prix, que ce soit à travers l'échange de permis ou de taxes⁵³. » Les énergies fossiles sont très peu coûteuses, raison pour laquelle nous avons besoin

d'institutions comme le plafonnement et l'échange ou une taxe carbone pour augmenter leur prix par un choix politique. Il nous faut viser le point d'intersection entre la courbe descendante des prix des technologies alternatives et la courbe ascendante des prix des combustibles fossiles, où ces derniers perdent leur avantage compétitif. Shue en conclut que « dès lors que cette intersection aura lieu, même l'égoïste pur et dur qui ne se soucie ni de l'environnement ni des droits d'autrui trouvera plus efficace d'utiliser des combustibles alternatifs⁵⁴. »

CONCLUSION

J'espère être parvenu à montrer qu'en utilisant certains mécanismes de marché, l'équité et l'efficacité environnementale peuvent être des objectifs conciliables. Même si une taxe carbone internationale n'est pas pour demain, et même si un marché global du carbone pose de nombreux problèmes qui n'ont pas été abordés ici, il est crucial de chercher à formuler des solutions justes et efficaces au changement climatique. Si la compréhension théorique du problème a déjà fait l'objet de nombreuses études philosophiques pertinentes, l'aspect pratique de sa résolution fait encore trop souvent défaut dans le champ de recherche de la justice climatique.

Bien sûr, certains éléments discutés ci-dessus sont largement débattus en éthique climatique, comme le critère de distribution des droits d'émission. Mais il me semblait que d'autres points méritaient d'être plus approfondis, notamment la justification de l'utilisation des incitations et leur mise en œuvre sous forme de différentes politiques climatiques. Dans les discussions philosophiques portant sur les solutions au changement climatique, cet instrument est soit ignoré, soit considéré comme intrinsèquement immoral. Je ne nie pas que dans d'autres domaines certaines incitations puissent être illégitimes; je défends simplement que dans le cas du changement climatique, elles sont le plus souvent un outil justifié et efficace.

Les incitations peuvent être conçues comme un découragement ou un encouragement à adopter un certain comportement: tant une taxe carbone qu'un système de plafonnement et d'échange visent à encourager les individus à changer leurs habitudes et favoriser des sources d'énergie moins polluantes, en augmentant le prix des combustibles fossiles et en réduisant les coûts des énergies renouvelables. Si les deux présentent l'avantage majeur de pousser les agents à réduire leurs émissions même s'ils n'ont pas encore pris conscience des devoirs qu'ils ont à le faire ou du caractère vertueux d'un tel comportement, les marchés du carbone ont l'avantage supplémentaire d'être politiquement plus réalistes. Des solutions au changement climatique devant être mises en œuvre de manière de plus en plus urgente, il est peut-être préférable de développer davantage les institutions les plus faisables. De ce point de vue, une question cruciale serait de savoir comment parvenir à améliorer les systèmes de plafonnement et d'échange actuels. Ce mécanisme ayant été jusqu'ici, il faut l'avouer, relativement inefficace, il s'agit de se demander si cela justifie son remplacement par d'autres solutions ou si d'autres manières de l'institutionnaliser ne s'annoncent pas plus prometteuses.

Il convient pour terminer de relever que la justice requiert plus que le suffisantisme qui a servi à guider les principales étapes de la mise en œuvre de la taxe et de la distribution des droits d'émission. Il s'agissait seulement de s'assurer que la réalisation de ces dispositifs s'accompagne d'une conception minimaliste de la justice : pour qu'ils puissent devenir pleinement justes, d'autres exigences doivent être remplies. Le point faible majeur du suffisantisme est qu'il est très difficile de fixer le seuil absolu de suffisance (correspond-il à une vie décente ou à une vie satisfaisante; et dans les deux cas, à quoi cela correspond-il concrètement?) et de justifier pourquoi, une fois ce seuil garanti pour tous, les inégalités ne doivent pas continuer à être réduites. Si les inégalités empêchant les individus de s'élever au-dessus du seuil sont certes les plus inévitables, il existe encore d'autres inégalités injustes une fois le seuil passé, raison pour laquelle le suffisantisme est souvent insuffisant. Entrer dans les débats théoriques complexes entre égalitarisme et suffisantisme nous aurait trop éloigné des ambitions de cet article; il est cependant important de les mentionner en conclusion pour montrer que beaucoup reste à faire avant de parvenir à une éthique climatique satisfaisante.

NOTES

- ¹ Pour leurs commentaires très utiles sur des versions antérieures de cet article, je tiens à remercier mes deux directeurs de thèse, Simone Zurbuchen et Alain Renaut, les commentateurs anonymes de la revue, ainsi que les personnes présentes lors du 7th *Uppsala Forum Workshop on Global Climate Change* le 8 avril 2013, notamment Edward Page et Aaron Maltais.
- ² Light, Andrew, “Finding a Future for Environmental Ethics,” *Les Ateliers de l'éthique/The Ethics Forum*, vol. 7, no. 3, 2012, pp. 71-80.
- ³ Un examinateur anonyme m’a proposé de remplacer « incitation » par « incitatif » pour traduire « *incentive* ». Je conserve la première traduction pour la simple raison qu’elle est plus souvent utilisée dans la littérature francophone sur le sujet. L’autre proposition est également tout à fait correcte, mais elle aurait créé un décalage avec le reste de la littérature, alors que mon objectif est de concilier l’approche économique, où ce type d’instrument est souvent discuté, avec une approche philosophique, ce qui implique d’utiliser autant que possible la même terminologie.
- ⁴ Voir par exemple : Pickering, Jonathan, Vanderheiden, Steve, et Seumas Miller, “‘If Equity’s In, We’re Out’: Scope for Fairness in the Next Global Climate Agreement,” *Ethics & International Affairs*, vol. 26, no. 4, 2012, pp. 423-443.
- ⁵ Je ne traite pas ici de la phase de justification, ce sujet très complexe nécessitant à lui seul un et même sans doute plusieurs articles : cette thématique est en effet l’une de celles qui a le plus fait couler d’encre en éthique climatique. Mon analyse se situe ici à un niveau différent par rapport à celui de l’enquête théorique cherchant à établir qui doit remplir les devoirs de justice climatique et pour quelles raisons : adoptant une approche plus pratique, je me demande plutôt quel type de politique pourrait permettre à ceux qui sont soumis aux principes de justice climatique de remplir partiellement leurs devoirs. Un article important permettant de se familiariser avec le premier niveau de débat est celui de Simon Caney, “Climate Change and the Duties of the Advantaged,” *Critical Review of International and Social Philosophy*, vol. 13, no. 1, 2010, pp. 203-228.
- ⁶ Gardiner, Stephen M., Caney, Simon, Jamieson, Dale et Henry Shue (eds.), *Climate Ethics: Essential Readings*, Oxford, Oxford University Press, 2010.
- ⁷ Jamieson, Dale, “Ethics, Public Policy, and Global Warming”, in Gardiner et al., *Climate Ethics: Essential Readings*, op. cit., pp. 77-86; pp. 79, 84.
- ⁸ Shue, Henry, “Subsistence Emissions and Luxury Emissions”, in Gardiner et al., *Climate Ethics: Essential Readings*, op. cit., pp. 200-214; p. 212.
- ⁹ Gardiner, Stephen M., “Ethics and Global Climate Change”, in Gardiner et al., *Climate Ethics: Essential Readings*, op. cit., pp. 3-35; pp. 3, 22.
- ¹⁰ Pour une bonne présentation des objectifs et du fonctionnement du GIEC, voir Ypersele (van), Jean-Pascal et Bruna Gaino, « Comment le GIEC gère-t-il les incertitudes scientifiques? », in Zaccà, Edwin, Gemenne, François et Jean-Michel Decroly (dir.), *Controverses climatiques, sciences et politique*, Paris, Les Presses de Sciences Po, 2012, pp. 77-96.
- ¹¹ Jamieson, “Ethics, Public Policy, and Global Warming”, op. cit., pp. 80-82. Heureusement, Jamieson est revenu sur cette position radicale, et écrit aujourd’hui que « [m]algré leurs limites, les modèles, réflexions et considérations économiques sont importantes et utiles. » Il continue en émettant quelques pages plus loin les réserves suivantes, qui sont tout à fait légitimes : « [I]es ressources de l’économie doivent jouer un rôle dans nos réflexions sur le changement climatique, mais nous devons faire preuve d’humilité épistémologique et normative. Nous devons être épistémologiquement humbles, parce que nous savons très peu du futur qui nous attend. Nous devons être normativement humbles, parce que l’analyse économique du changement climatique repose sur des présupposés normatifs qu’elle n’est pas en mesure de justifier. » (Jamieson, Dale, *Reason in a Dark Time: Why the Struggle Against Climate Change Failed – and What it Means for our Future*, Oxford/New York, Oxford University Press, 2014, pp. 137, 143).
- ¹² Voir notamment : Dietz, Simon, Hepburn, Cameron et Nicholas Stern, “Economics, Ethics and Climate Change”, in Basu, Kaushika et Ravi Kanbur (eds), *Arguments for a Better World*:

Essays in Honour of Amartya Sen. Volume II: Society, Institutions, and Development, Oxford, Oxford University Press, 2008, pp. 365-386.

- ¹³ Caney, Simon et Cameron Hepburn, “Carbon Trading: Unethical, Unjust and Ineffective?,” *Royal Institute of Philosophy Supplement*, vol. 69, 2011, pp. 201-234.
- ¹⁴ Broome, John, *Climate Matters: Ethics in a Warming World*, New York/London, W. W. Norton & Company, 2012.
- ¹⁵ Dion, Stéphane, « La lutte contre les changements climatiques : pourquoi le Canada en fait-il si peu? », *La Revue Tocqueville*, vol. 32, no. 2, 2011, pp. 21-49; p. 26.
- ¹⁶ Helm, Dieter, *The Carbon Crunch: How We're Getting Climate Change Wrong – and How to Fix It*, New Haven/London, Yale University Press, 2012, p. 42.
- ¹⁷ Archer, David, *The Long Thaw: How Humans are Changing the next 100,000 Years of Earth's Climate*, Princeton, Princeton University Press, 2009, p. 11.
- ¹⁸ Gardiner, Stephen M., *A Perfect Moral Storm: The Ethical Tragedy of Climate Change*, Oxford, Oxford University Press, 2011, pp. 24-38; pp. 108-123. Gardiner reprend sa version complexifiée de la tragédie des communs dans un article qui paraîtra prochainement en vue de montrer ses implications au niveau institutionnel: Gardiner, Stephen M., “A Call for a Global Constitutional Convention Focused on Future Generations,” *Ethics & International Affairs*, à paraître.
- ¹⁹ Dion, Stéphane et Éloi Laurent, « De Rio à Rio : Un signal-prix mondial du carbone pour sortir de la grande incohérence climatique », in *Documents de travail OFCE*, no. 15, 2012, pp. 1-19; p. 13 (<http://www.ofce.sciencespo.fr/pdf/dtravail/WP2012-15.pdf>).
- ²⁰ Pour un graphique illustrant très bien cette réalité, voir Stern, Nicholas, “The Economics of Climate Change,” in Gardiner *et al.*, *Climate Ethics: Essential Readings*, *op. cit.*, pp. 39-76; p. 64.
- ²¹ Voir par exemple : Goodin, Robert E., “Selling Environmental Indulgences,” in Gardiner *et al.*, *Climate Ethics: Essential Readings*, *op. cit.*, pp. 231-246.
- ²² Grant, Ruth W., *Strings Attached: Untangling the Ethics of Incentives*, Princeton, Princeton University Press, 2012.
- ²³ Ces mesures proposées par des écologistes américains radicaux sont présentées et critiquées par Simon, Julian, *The Ultimate Resource 2*, Princeton, Princeton University Press, 1996, p. 569.
- ²⁴ Jamieson, Dale, “Energy, Ethics, and the Transformation of Nature,” in Arnold, Denis G. (ed.), *The Ethics of Global Climate Change*, Cambridge, Cambridge University Press, 2011, pp. 16-37 ; p. 36.
- ²⁵ Szasz, Andrew, “Is Green Consumption Part of the Solution?,” in Dryzek, John S., Norgaard, Richard B. et David Schlosberg (eds.), *The Oxford Handbook of Climate Change and Society*, Oxford, Oxford University Press, 2011, pp. 594-608.
- ²⁶ Light, “Finding a Future for Environmental Ethics”, *op. cit.*, p. 75.
- ²⁷ Valentini, Laura, “Cosmopolitan Justice and Rightful Enforceability,” in Brock, Gillian (ed.), *Cosmopolitanism Versus Non-Cosmopolitanism: Critiques, Defenses, Reconceptualizations*, Oxford, Oxford University Press, 2013, pp. 92-110 ; p. 95.
- ²⁸ Certaines options supplémentaires que les individus ont à leur disposition pour réduire spontanément leurs émissions sont développées dans Dietz, Thomas, Gardner, Gerald T., Gilligan, Jonathan, Stern, Paul C. et Michael Vandenberg, “Household Actions Can Provide a Behavioral Wedge to Rapidly Reduce US Carbon Emissions,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 44, 2009, pp. 18452-18456. Pour une justification morale du devoir de réduire ses émissions individuelles, voir : Schwenkenbecher, Anne, “Is There an Obligation to Reduce One's Individual Carbon Footprint?,” *Critical Review of International Social and Political Philosophy*, vol. 17, no. 2, 2014, pp. 168-188.
- ²⁹ Laurent, Éloi et Jacques Le Cacheux, « Carbone sans frontières. Quelles solutions fiscales face aux émissions importées? », *Revue de l'OFCE*, vol. 122, 2012, pp. 83-98 ; p. 88.
- ³⁰ Stern, “The Economics of Climate Change”, *op. cit.*, p. 39.

- ³¹ PNUD, *Rapport mondial sur le développement humain 2007-2008. La lutte contre le changement climatique. Un impératif de solidarité humaine dans un monde divisé*, Paris, La Découverte, 2008, p. 11.
- ³² Pour une manière ingénieuse de mettre en œuvre le principe du pollueur payeur ainsi que celui de la capacité à payer, qui pourrait aider à déterminer le montant de la taxe pour chaque pays, voir Baer, Paul, “The Greenhouse Development Right Framework for Global Burden Sharing: Reflection on Principles and Prospects,” *WIREs Climate Change*, vol. 4, no. 1, 2013, pp. 61-71.
- ³³ Dion, Laurent, « De Rio à Rio », *op. cit.*, p. 13.
- ³⁴ *Ibid.*
- ³⁵ Cohen, Steven, *Sustainability Management: Lessons from and for New York City, America, and the Planet*, New York, Columbia University Press, 2011.
- ³⁶ Je remercie le commentateur anonyme qui m’a proposé d’ajouter cet argument.
- ³⁷ Le changement climatique est en effet un problème de sécurité au même titre que les menaces posées par le terrorisme, les épidémies globales ou la prolifération nucléaire, raison pour laquelle le budget national militaire, notamment celui des États-Unis (qui équivaut aux dépenses militaires combinées des autres pays du monde), peut être considéré comme une source potentielle de support pour la lutte contre le changement climatique. Sur ce point, voir Sachs, Jeffrey, *Common Wealth: Economics for a Crowded Planet*, New York, The Penguin Press, 2008, pp. 281-283. Cette option est également envisagée par Henry Shue : « l’équivalent d’une année de budget pour l’occupation en Irak permettrait de mettre un terme à la pauvreté chronique et de financer plusieurs années d’action vigoureuse contre le changement climatique. » (Shue, Henry, “Deadly Delays, Saving Opportunities: Creating a More Dangerous World?,” in Gardiner *et al.*, *Climate Ethics: Essential Readings*, *op. cit.*, pp. 146-162; p. 161, note 34). Enfin, elle est également justifiée par Andrew Guzman, qui montre que les effets du changement climatique menacent la sécurité nationale des pays développés, notamment des États-Unis. Un exemple parmi beaucoup d’autres est celui des conséquences du réchauffement global sur le terrorisme : « la réduction de l’approvisionnement en eau, la baisse des rendements agricoles, l’augmentation du niveau des océans, ainsi qu’un climat plus chaud, vont contribuer à stimuler les tensions ethniques et religieuses existantes, intensifier la compétition pour les ressources, menacer la production et la distribution du pétrole, permettant ainsi au terrorisme de se répandre davantage. » (Guzman, Andrew T., *Overheated: The Human Cost of Climate Change*, Oxford, Oxford University Press, 2013, p. 170).
- ³⁸ Voir sur ce point l’analyse que Daniel Weinstock fait de la proposition de Thomas Pogge : Weinstock, Daniel, « L’éthique de la santé globale », in Jeangène Vilmer, Jean-Baptiste et Ryoa Chung (dir.), *Éthique des relations internationales. Problématiques contemporaines*, Paris, Presses universitaires de France, 2013, pp. 423-445.
- ³⁹ Sachs, *Common Wealth: Economics for a Crowded Planet*, *op. cit.*, pp. 34-35.
- ⁴⁰ Je remercie l’examinateur anonyme qui a attiré mon attention sur ce point : vu que, comme il le relève, 40 % des gaz à effet de serre produits au Québec proviennent du secteur des transports qui dépend à 99 % du pétrole, des mesures compensatrices dans ce secteur, ici et ailleurs, sont indispensables pour s’assurer que la taxe carbone soit équitable.
- ⁴¹ Helm, *The Carbon Crunch: How We’re Getting Climate Change Wrong – and How to Fix It*, *op. cit.*, p. 192.
- ⁴² Hoffmann, Matthew J., *Climate Governance at the Crossroads: Experimenting with a Global Response after Kyoto*, Oxford, Oxford University Press, 2011.
- ⁴³ Page, Edward A., “Cosmopolitanism, Climate Change, and Greenhouse Emissions Trading,” *International Theory*, vol. 3, no. 1, 2011, pp. 37-69 ; pp. 42-43.
- ⁴⁴ Hepburn, Cameron, “Carbon Taxes, Emissions Trading, and Hybrid Schemes,” in Helm, Dieter et Cameron Hepburn (eds.), *The Economics and Politics of Climate Change*, Oxford/New York, Oxford University Press, 2009, pp. 365-384 ; p. 381.
- ⁴⁵ Pour une bonne réponse à ces objections, voir notamment Caney, Simon, “Markets, Morality and Climate Change: What, if Anything, is Wrong with Emissions Trading?,” *New Political Economy*, vol. 15, no. 2, 2010, pp. 197-224.

- ⁴⁶ Dobson, Andrew, *Citizenship and the Environment*, Oxford, Oxford University Press, 2003, pp. 1-4.
- ⁴⁷ Spash, Clive L., "The Brave New World of Carbon Trading," *New Political Economy*, vol. 15, no. 2, 2010, pp. 169-196 ; pp. 189-190.
- ⁴⁸ Page, "Cosmopolitanism, Climate Change, and Greenhouse Emissions Trading," *op. cit.*, p. 52.
- ⁴⁹ À nouveau, il ne s'agit pas de remplacer un type d'éthique par un autre ; je cherche simplement à compléter l'éthique déontologique et celle des vertus par un autre type d'éthique plus efficace. La promotion de comportements moraux ou vertueux est une solution importante, mais insuffisante.
- ⁵⁰ Frankfurt, Harry, "Equality as a Moral Ideal," *Ethics*, vol. 98, no. 1, 1987, pp. 21-43.
- ⁵¹ Caney, Simon, "Just Emissions," *Philosophy & Public Affairs*, vol. 40, no. 4, 2012, pp. 255-300; p. 297.
- ⁵² Shue, Henry, "Human Rights, Climate Change, and the Trillionth Ton," in Arnold, *The Ethics of Global Climate Change*, *op. cit.*, pp. 292-314 ; pp. 310-311.
- ⁵³ Shue, Henry, "Climate Hope: Implementing the Exit Strategy," *Chicago Journal of International Law*, vol. 13, no. 2, 2013, pp. 381-402; p. 381. Toutes les publications de Henry Shue sur le changement climatique sont désormais rassemblées dans un recueil d'articles: Shue, Henry, *Climate Justice: Vulnerability and Protection*, Oxford, Oxford University Press, 2014.
- ⁵⁴ Shue, "Human Rights, Climate Change, and the Trillionth Ton", *op. cit.*, pp. 313-314.

EXPLOITATION ET OBLIGATION DE TRAVAILLER

PIERRE-ÉTIENNE VANDAMME

ASPIRANT FNRS, UNIVERSITÉ CATHOLIQUE DE LOUVAIN

RÉSUMÉ :

Cet article défend une définition de l'exploitation, restreinte aux relations de travail, en tentant d'une part d'explicitier une certaine compréhension de sens commun du concept (rémunération inéquitable en fonction du travail presté), et d'autre part d'échapper aux difficultés qui ont affecté la définition marxiste traditionnelle de l'exploitation comme extorsion de la plus-value (dans ses diverses variantes). Il explore ainsi le lien entre l'exploitation et l'obligation matérielle de travailler pour subvenir à ses besoins fondamentaux. Après avoir mis en garde contre les politiques d'activation des chômeurs, il conclut que l'exploitation est un phénomène contre lequel on peut lutter à l'aide de mécanismes relativement simples, même dans les sociétés capitalistes. Il rappelle toutefois que cela ne suffit pas à réaliser la justice sociale, resituant l'exploitation parmi d'autres enjeux fondamentaux pour une philosophie politique égalitariste

ABSTRACT:

This paper offers a definition of exploitation, restricted to work relationships, that pretends to be closer to a common sense understanding of the concept – an unfair remuneration in relation to the work performed. The proposed definition also aims at avoiding the troubles that have affected the traditional Marxist definition of exploitation as surplus value extraction in its various forms. Thus, the paper investigates the connection between exploitation and the material obligation to work in order to satisfy one's basic needs. It warns against the damages induced by the politics of activation of the unemployed. And it concludes that exploitation can be fought using relatively simple instruments, even in capitalist societies. Nonetheless, relocating exploitation among other fundamental concerns for an egalitarian political philosophy, it reminds that the fight against exploitation is not enough for the achievement of social justice.

INTRODUCTION

Parmi les travaux des dernières décennies sur le concept marxiste d'exploitation, des avancées particulièrement intéressantes ont été effectuées par les approches analytiques du marxisme (en particulier celles de G. A. Cohen, John Roemer, Jon Elster, Philippe Van Parijs et Richard Arneson). Peu à peu, la conception marxiste traditionnelle de l'exploitation comme extorsion de la plus-value a été abandonnée, principalement en raison de l'hypothèse libertarienne de propriété de soi de laquelle elle semblait difficilement dissociable. Le problème de l'exploitation a dès lors été jugé non essentiel (Cohen, 1995), ou déplacé vers les questions de propriété des moyens de production (Roemer, 1982), ou élargi à d'autres injustices touchant notamment les étrangers et les chômeurs (Van Parijs, 1993), perdant alors quelque peu de sa spécificité.

Partant, il me semble qu'une certaine compréhension partagée de la notion d'exploitation a été perdue : celle que nous invoquons au quotidien pour qualifier une relation injuste entre le travail d'un individu et sa rémunération. Je suggère dans ce qui suit que cette compréhension de sens commun de l'exploitation ne repose pas sur un sentiment de propriété de soi ou une réclamation d'un droit à l'entièreté des fruits de son travail, comme dans la conception marxiste traditionnelle, ou à la propriété collective des moyens de production, comme chez Roemer, mais qu'elle naît plutôt de la relation entre la liberté d'accepter un travail, le pouvoir de négociation des conditions salariales et la rémunération obtenue. Ce faisant, je ne démens pas le fait qu'on puisse utiliser avec plus ou moins de succès la notion d'exploitation en dehors des relations de travail, mais je me focalise ici sur ces dernières, auxquelles le concept me semble intimement lié. Avant d'explicitier la définition de l'exploitation que j'entends défendre ici et d'explorer ses implications, je commencerai par rappeler quelques difficultés auxquelles sont confrontées deux conceptions marxistes de l'exploitation : l'une traditionnelle, proposée par Marx et défendue par plusieurs de ses disciples ; l'autre proposée plus récemment par John Roemer.

L'EXTORSION DE LA PLUS-VALUE

Selon Marx et certains marxistes, les relations de travail, en régime capitaliste, sont nécessairement (et de manière évidente) injustes dès lors que l'employeur tire une plus-value du travail presté par l'employé. Sans cette plus-value, il n'aurait d'ailleurs aucun intérêt à engager ce travailleur, puisque les entreprises opèrent généralement, dans ce contexte capitaliste, dans une logique de maximisation du profit personnel. En conséquence, le travailleur se voit dépossédé d'une part de la valeur liée à son travail (la plus-value). Il est donc, de ce fait, exploité (c'est un vol, une extorsion). Présenté de manière analytique, l'argument cohérent (Cohen, 1988, p. 228 pour la formulation ; Kymlicka, 2002, p. 179 pour la présentation)¹ est le suivant :

1. Le travailleur est la personne qui crée le produit, lequel a de la valeur.
2. Le capitaliste s'approprie une part de la valeur du produit.
Donc :
3. Le travailleur reçoit moins de valeur que la valeur de ce qu'il crée.

4. Le capitaliste s'approprié une part de la valeur de ce que crée le travailleur.

Donc :

5. Le travailleur est exploité par le capitaliste.

Pour peu que l'on reconnaisse la connotation morale négative du concept d'exploitation, cet argument implique la prémisse selon laquelle il est injuste que le travailleur soit privé d'une part des fruits de son travail (ou qu'on s'approprié une part de son temps de travail, dans une autre formulation possible). Or, cette prémisse est douteuse, car elle semble reposer sur le principe de propriété de soi (Cohen, 1995, pp. 146-147), dont Robert Nozick (1974) a fait un brillant usage pour défendre un libéralisme de droite et un État minimal – ce qui place les marxistes en bien mauvaise compagnie, eu égard à leur critique du capitalisme. En effet, le droit de chacun à l'ensemble des fruits de son travail semble exclure, outre l'extorsion de la plus-value par le capitaliste, également toute forme de taxe gouvernementale sur le travail (pour compenser, par exemple, l'absence de revenus liés au travail pour certaines personnes).

Une première tentative pour échapper à cette conclusion indésirable d'un point de vue marxiste consiste à dire, comme le fait Elster (1978, p. 3), qu'il n'y a exploitation dans un sens marxiste compréhensif qu'à condition que la différence entre ce que le travailleur produit et ce qu'il reçoit ne puisse être justifiée par une redistribution selon les besoins. Cependant, la conception marxiste traditionnelle de l'exploitation – même modérée par la clause d'Elster – a des implications très inégalitaires, particulièrement fortes dans un contexte capitaliste où ce sont les gains individuels qui priment. En effet, les variations de talents entraînent des variations dans la capacité individuelle de gagner sa vie grâce à son travail, donc des inégalités *injustifiables* de revenu. Puisqu'un manque de talent et un handicap font partie des « circonstances » desquelles un individu ne peut raisonnablement être tenu responsable (Rawls, 1971 ; Dworkin, 1981b ; Cohen, 1989 ; Roemer, 1996 ; Spitz, 2008), un État tolérant ces inégalités est injuste, faisant preuve de complaisance à l'égard des injustices naturelles. Ce qui implique des redistributions qui vont au-delà des « besoins » évoqués par Marx et Elster.

Une autre manière d'essayer de sauver, malgré cela, l'idée d'un droit aux fruits du travail, est d'en faire un droit collectif : l'ensemble des travailleurs a droit aux bénéfices de son travail collectif, qu'il redistribuera en son sein comme juste lui semble. Mais demeurent alors une série de problèmes, comme le statut des chômeurs et de ceux qui sont dans l'incapacité de travailler ; les inégalités de capacités productives collectives entre différentes entreprises ; ou encore (peut-être la plus grande source d'injustices) les inégalités de revenu pour un même travail entre contextes nationaux distincts. Dès lors, si les marxistes admettent l'idée selon laquelle des inégalités de revenu liées à la prédisposition génétique ou la situation géographique, notamment, sont injustifiables – et je ne doute pas qu'ils l'admettent –, ils doivent défendre diverses redistributions des fruits variables du travail, donc abandonner leur conception traditionnelle de l'exploitation.

Si les implications antiredistributives indésirables de cette dernière ont pu échapper pendant longtemps aux marxistes, c'est sans doute en raison de la relative congruence passagère entre la catégorie des plus démunis et celle des travailleurs (Cohen, 1995, pp. 152-158). Mais dans le contexte contemporain, la « classe » des plus démunis concerne avant tout les chômeurs, les personnes âgées, handicapées (et bien entendu, les travailleurs précaires), personnes vulnérables qui ne peuvent vivre que de redistributions continues du produit social (donc de taxes sur le travail en plus du capital). Si bien que le concept d'exploitation doit être repensé si l'on veut éviter qu'il soit vulnérable à l'idée détestable selon laquelle les travailleurs sont nécessairement exploités lorsque sont taxés les revenus du travail.

Plus fondamentalement, c'est l'idée de propriété de soi – fondement vraisemblablement insoupçonné du concept marxiste traditionnel d'exploitation – qui doit être réexaminée. Or, elle s'avère à l'examen très contestable, comme l'a brillamment défendu G. A. Cohen². D'abord, parce que contrairement à ce qu'avancent certains de ses défenseurs, renoncer au principe de propriété de soi n'implique pas d'accepter l'esclavage, c'est-à-dire le droit d'être propriétaire d'autres personnes. La liberté individuelle peut être défendue sans que l'on se réfère à cette propriété de soi – et elle l'est peut-être même mieux ainsi. En effet, il se peut que des situations de quasi-esclavage émergent plus facilement dans un monde (nozickéen) qui respecterait la propriété de soi, puisqu'il y est plus difficile de contester la validité morale des contrats passés entre travailleurs indigents et capitalistes opulents, par exemple. Ensuite, contrairement à ce que défend Nozick, le principe de propriété de soi n'offre pas davantage d'autonomie aux individus. Il les rend plutôt prisonniers de leurs « circonstances » (leurs talents), qui (en relation avec celles des autres) déterminent leur niveau de vie et leurs opportunités. Enfin, rejeter la propriété de soi n'implique pas non plus de rejeter la maxime kantienne selon laquelle il ne faut jamais traiter autrui *seulement comme un moyen* mais toujours *aussi comme une fin*³. Exiger des personnes talentueuses qu'elles travaillent davantage que le strict nécessaire suffisant à combler leurs besoins et préférences, cela afin de redistribuer aux moins talentueux une part des richesses produites, c'est certes traiter les talentueux comme des moyens (de réaliser la justice sociale). Mais cela n'empêche aucunement de les traiter aussi comme des fins (pour peu que l'on puisse clarifier ce concept), c'est-à-dire comme des personnes également dignes de respect et d'attention. Au vu de ces arguments, les marxistes ne doivent pas s'effrayer à l'idée d'abandonner le principe de propriété de soi⁴ (et il se peut même que ce principe perde de l'attrait aux yeux de nombreux libertariens, suggère Cohen).

Cependant, même si l'on devait admettre le principe de propriété de soi, l'idée selon laquelle il est injuste que le travailleur soit privé d'une part des fruits de son travail ou qu'on s'approprié une part de son temps de travail demeurerait douteuse, parce que l'idée de propriété de soi n'implique pas logiquement la propriété des fruits de son travail, au contraire de ce que pensait Locke (1690), premier philosophe à avoir évoqué ce double principe. En effet, le travailleur fait généralement usage, dans son travail, d'instruments et de ressources dont il

n'est pas l'unique propriétaire. Les instruments peuvent avoir été construits par d'autres et achetés par l'employeur (investissement sur lequel il peut légitimement réclamer un retour). Quant aux ressources naturelles utilisées, à qui appartiennent-elles? Juridiquement parlant, à personne. Mais moralement parlant, à l'ensemble de la communauté, dès lors qu'il n'y a aucune raison morale de se satisfaire d'un principe aussi vulgaire que celui du « premier arrivé, premier servi » (Van Parijs, 1993 ; Cohen, 1995). Il ne saurait dès lors être question d'un droit légitime du travailleur à l'ensemble des fruits de son travail.

Considérant ces diverses difficultés, il a semblé à certains que si le travailleur s'engageait *volontairement* dans une relation de travail avec un capitaliste, conscient qu'il ne posséderait pas l'entièreté de la valeur liée à son travail, l'exploitation paraissait moins évidente. C'est pourquoi, explique Will Kymlicka (2002, p. 179), la plupart des marxistes qui défendent la définition traditionnelle de l'exploitation ont ajouté une clause selon laquelle une situation est injuste uniquement si le travailleur est forcé de travailler pour le capitaliste⁵ – ce qui modifie largement la compréhension traditionnelle de l'exploitation et la rapproche de la conception que je défendrai à partir de la section 3.

LA PROPRIÉTÉ DES MOYENS DE PRODUCTION

Pour remédier à cette faiblesse de la définition de l'exploitation comme extorsion de la plus-value, John Roemer déplace la question de l'exploitation capitaliste de la relation travailleur-capitaliste vers la distribution sociétale des moyens de production. Des personnes seraient victimes de l'exploitation capitaliste, dit-il, à condition que leur sort puisse être amélioré si les moyens de production venaient à être possédés par la collectivité (Roemer, 1982, pp. 202-211). Cette nouvelle définition de l'exploitation capitaliste permet d'intégrer les laissés pour compte de la formulation traditionnelle de l'exploitation, à savoir les chômeurs involontaires, les personnes âgées, les travailleurs domestiques et les personnes handicapées, notamment. Autre avantage : elle permet d'éviter les impasses redistributives de la conception traditionnelle. L'appropriation d'une part de la valeur liée au travail n'est pas forcément injuste. Elle est injuste à la seule condition qu'elle laisse le travailleur dans une situation plus défavorable que s'il était également propriétaire des moyens de production (mais ce n'est pas forcément le cas, puisque le capitaliste peut, pour diverses raisons imaginables, être dans une position de négociation défavorable au moment d'établir le contrat de travail). Au contraire, cette appropriation est juste, par exemple, si elle permet à l'État de compenser les inégalités liées aux circonstances des individus (position sociale et talents), pour lesquelles ils ne peuvent être tenus responsables et au nom desquelles rien ne justifie qu'ils soient systématiquement désavantagés.

Ce déplacement de la notion d'exploitation par Roemer est habile, puisqu'il permet de sortir d'une impasse vers laquelle menait la conception marxiste traditionnelle de l'exploitation, tout en restant fidèle à l'intuition marxiste selon laquelle c'est dans le régime de propriété des moyens de production que réside la cause de l'exploitation. Cependant, la notion d'exploitation devient très générale ; elle ne concerne plus la relation de travail, mais l'ensemble des positions

dans un régime économique donné. C'est ce qui permet à Van Parijs (1993, pp. 110-152) d'élargir la notion d'exploitation aux relations entre travailleurs et chômeurs, ainsi qu'entre nationaux et migrants. Or, le problème de la conception traditionnelle de l'exploitation n'était pas qu'elle se focalisait sur les travailleurs, mais que le principe normatif qu'elle défendait avait des implications négatives sur les plus démunis non-travailleurs. Cela étant posé, il me semble intéressant de distinguer les injustices liées aux différences de pouvoir de négociation des contrats de travail (exploitation) de celles liées aux variations d'opportunités engendrées par l'arbitraire génétique et géographique de la naissance. En effet, la lutte contre ces différentes injustices peut appeler à chaque fois des solutions différentes.

Cohen (1995, p. 151) mobilise également la très large définition standard de l'exploitation : exploiter une personne, c'est profiter injustement (*taking unfair advantage*) d'elle. Dans un cas (Roemer) comme dans l'autre (Cohen), il me semble que beaucoup trop de relations tombent sous ces définitions pour que le concept d'exploitation préserve sa spécificité (caractériser une injustice propre aux relations de travail) et son intérêt (offrir une compréhension plus fine des différentes dimensions des injustices sociales).

D'ailleurs, dans la prolongation des travaux de Cohen et Roemer, notamment, la préoccupation des philosophes politiques analytiques d'influence marxiste s'est déplacée vers des problèmes plus généraux de justice distributive⁶ dans lesquels la notion d'exploitation ne joue plus du tout un rôle central. Ce qui importe, désormais, aux yeux de Cohen (1989), Arneson (1989), Van Parijs (1995) et Roemer (1998), c'est la distribution des ressources et l'égalité des chances.

Que reste-t-il alors de l'intuition largement répandue selon laquelle certains travailleurs sont exploités en raison d'une rémunération non équitable? Quand un individu affirme qu'il est exploité par son employeur, il ne pense pas forcément qu'il devrait avoir droit à l'ensemble des fruits de son travail, car il retirerait très probablement sa plainte s'il obtenait un autre niveau de revenu (quoi qu'il en soit de la plus-value obtenue par l'employeur). Il n'incrimine pas non plus nécessairement la propriété privée des moyens de production. Et l'on ne songe pas non plus à celle-ci lorsqu'on juge qu'un grand nombre de travailleurs chinois, par exemple, sont aujourd'hui exploités.

Si l'exploitation, selon Marx, est forcément liée au régime de propriété privée des moyens de production, c'est parce qu'il n'imagine pas du tout un régime d'État-providence au sein duquel les travailleurs pourraient ne pas être propriétaires des moyens de production et pourtant ne pas être forcés à travailler, ni exploités. Dans la perspective de Marx, puisque les travailleurs ne possèdent pas de moyens de production, ils sont incapables d'assurer leur subsistance sans s'engager dans des contrats de travail iniques avec des capitalistes (Arneson, 1981, pp. 225-226). Mais si l'État joue un rôle régulateur fort, encadrant les relations de marché, les perspectives des travailleurs sont susceptibles de changer ; ils peuvent échapper à l'alternative entre être exploités et mourir de faim.

L'exploitation n'est donc pas propre aux régimes capitalistes – elle le serait plutôt aux régimes injustes. En effet, le sentiment d'exploitation pourrait disparaître dans un régime capitaliste pour peu que chacun dispose du pouvoir d'obtenir une rémunération équitable pour son travail (ce qui ne signifie pas que toutes les injustices auraient disparu, puisque l'exploitation n'est qu'une dimension précise des injustices sociales, celle qui concerne le triangle employé-salaire-employeur).

EXPLOITATION ET DÉPENDANCE PAR RAPPORT AUX BESOINS FONDAMENTAUX

Il me semble que l'on peut chercher à expliciter la compréhension de sens commun de l'exploitation en définissant cette dernière de la manière suivante : *une personne est exploitée si le salaire qu'elle obtient en échange de son travail est inférieur à ce qu'il serait si elle bénéficiait d'une garantie inconditionnelle de voir ses besoins fondamentaux couverts par l'État*. En effet, c'est la nécessité de pourvoir (à tout prix) à leurs besoins fondamentaux qui fait accepter aux travailleurs des salaires iniques. Or, l'existence d'un droit inconditionnel à un revenu suffisant pour couvrir ces besoins leur octroierait un véritable pouvoir de négociation relativement à l'employeur et à la détermination du salaire (la liberté réelle de refuser un emploi ou d'y renoncer dans des conditions jugées comme relevant de l'exploitation). Les salaires reflèteraient alors de manière plus adéquate le *coût subjectif du travail* (rapport de l'individu aux tâches à réaliser et aux conditions de travail), celui-là même qui nourrit nos jugements sur l'exploitation.

À cette définition de l'exploitation, j'ajoute deux clauses. 1) La personne que l'on juge exploitée ou non ne dispose pas par ailleurs de revenus liés au capital. Cette précision assez évidente doit permettre de contrer l'éventuel contre-exemple du riche rentier qui refuserait de travailler pour moins de 50€/heure, même dans un emploi extrêmement peu pénible. Ce type de préférence dispendieuse n'est possible que grâce aux revenus assurés par son capital. 2) La personne en question n'est pas en possession d'un talent tellement rare ou valorisé sur le marché de l'emploi qu'il lui confère nécessairement un pouvoir de négociation favorable. On le voit, ma définition de l'exploitation est liée à une *structure de négociation* – ce qui la rapproche quelque peu des conceptions de l'exploitation inspirées de la théorie des jeux (Elster, 1978 ; Roemer, 1982 ; Barry, 1989).

Maintenant, pourquoi ne pas simplement dire qu'un travailleur n'est pas exploité tant qu'il jouit de la liberté de travailler ou non? Parce que la « liberté » fait partie des concepts les plus flous qui puissent être utilisés. La droite défend le libre marché au nom de la liberté ; la gauche, la régulation en ce même nom. Certains libertariens pensent que les relations de travail entre prolétaires et capitalistes ne sont pas problématiques dès lors que les prolétaires sont libres de signer ou non les contrats de travail qui les unissent aux capitalistes. Les marxistes, eux, considèrent les contrats comme des « impostures de liberté » (Cohen, 1995, p. 151). La liberté, tant qu'elle demeure indéfinie, est un concept vide, et donc dangereux. C'est pourquoi, dans ma définition de l'exploitation, je précise en

quoi consiste la liberté réelle de ne pas travailler : l'*indépendance par rapport à la satisfaction des besoins fondamentaux*.

Examinons, avec cette nouvelle définition de l'exploitation, l'exemple suivant, emprunté à Cohen (1995, pp. 149-150) :

Un travailleur qui apprécie son travail et le salaire qu'il en obtient est employé par son voisin infirme, qui mène une vie misérable mais, au contraire du travailleur, possède des moyens de production. Ce capitaliste infirme s'approprie, des fruits du travail de son employé, tout juste de quoi se maintenir en vie⁷.

Cette relation de travail relève-t-elle de l'exploitation? Oui, selon la définition marxiste traditionnelle (même si ce n'est certainement pas l'intention de ceux qui la mobilisent d'en arriver à une telle conclusion). Pas forcément, selon la définition de Roemer, car il n'est pas certain que l'employé bénéficierait d'une appropriation collective des moyens de production. En effet, il devrait quand même indemniser le capitaliste pour son incapacité à travailler, ou l'ex-capitaliste deviendrait lui-même victime d'exploitation socialiste, cette exploitation des moins talentueux par les talentueux qui demeure une fois effectuée l'égalisation des ressources externes (Roemer, 1982, ch. 8).

Qu'en est-il selon ma définition de l'exploitation? Tout dépend de la structure d'interdépendance entre les deux protagonistes. Si le travailleur possède la liberté réelle de renoncer à son travail, il ne doit pas être considéré comme exploité. Admettons maintenant que, dans l'exemple de Cohen, il n'y ait pas de droit inconditionnel au revenu (ni pour le travailleur, ni pour le capitaliste infirme). Il demeure peu plausible que le travailleur soit exploité alors que son employeur a impérativement besoin d'un employé pour se maintenir lui-même en vie. L'employeur possède la position de négociation la plus défavorable (étant donné son infirmité et malgré son capital), ce qui fait qu'il n'est même pas en mesure d'essayer d'exploiter son employé. Il se pourrait même qu'il le surpaye, c'est-à-dire que le travailleur profite de sa faiblesse pour imposer des conditions salariales qui désavantagent l'employeur en comparaison avec une situation d'égal pouvoir de négociation.

Est-ce alors le capitaliste qui est exploité? Si l'on répond oui à cette question, le risque est d'en revenir à assimiler l'exploitation au fait de profiter de quelqu'un, ce qui me paraît trop général. Ce qu'il faut plutôt répondre, me semble-t-il, c'est que d'autres injustices doivent être considérées, outre l'exploitation ; en l'occurrence, une inégalité de revenus moralement injustifiable, car dérivant d'une distribution inégale des ressources « internes » (talents et capacités).

BESOINS FONDAMENTAUX ET PRÉFÉRENCES PERSONNELLES

Revenons à la proposition selon laquelle l'exploitation se mesure à l'aune de la dépendance du travailleur à l'égard de la satisfaction de ses besoins fondamentaux. Pour éviter une première salve d'objections, il convient de préciser que la

notion de « besoins fondamentaux » d'un individu doit inclure les besoins des personnes à sa charge (sans quoi elle est factice). Si cette notion demeure cependant floue, c'est parce qu'elle est évidemment relative aux contextes sociaux et au coût de la vie. Elle est en outre relative aux différentes caractéristiques physiques et cognitives individuelles, et sa perception par les acteurs variera en fonction de leurs préférences personnelles. Imaginons que l'État décide de couvrir l'ensemble des besoins fondamentaux des citoyens. Doit-il prendre en considération ces variations interpersonnelles?

Il a été remarqué qu'il serait vraisemblablement injuste que l'ensemble de la communauté doive assumer le coût de préférences individuelles dispendieuses et qu'il serait dès lors justifié que chacun assume le coût de ses préférences (Dworkin, 1981a), quitte à les revoir à la baisse en fonction du revenu disponible. Par contre, il paraît tout aussi injuste de ne pas considérer les divergences de besoins liées à des circonstances subies (maladies, accidents, handicaps notamment). C'est un point qui a notamment été défendu avec brio par Amartya Sen dans son approche dite des « capacités » (Sen, 1992 ; 2009), et qui s'accorde parfaitement avec le principe marxiste « À chacun selon ses besoins ». Mais il a également été souligné que les préférences individuelles, dans une certaine mesure, peuvent elles aussi être « subies », et qu'il est extrêmement difficile d'établir une stricte dichotomie entre les besoins et les préférences (Arneson, 1989 ; Cohen, 1989 ; Spitz, 2008).

Comment l'État pourrait-il alors s'y prendre pour couvrir les besoins fondamentaux d'une manière qui soit relativement insensible aux choix, mais sensible aux divergences de caractéristiques internes? Il ne semble pas exister de formule idéale pour institutionnaliser l'approche par les capacités, et sans doute encore moins celles d'Arneson et Cohen, qui visent l'égalité des chances de bien-être, celles-ci requérant une quantité d'informations sur les individus que l'État n'obtiendra jamais, et une complexité distributive inimaginable. Sen (2009) ne semble pas s'en formaliser, lui qui juge plus utile de fournir des outils permettant d'évaluer de manière comparative plusieurs états sociaux que de formuler un idéal de société parfaitement juste. Quant à Arneson et Cohen, ils demeurent relativement discrets sur cette question.

Sans doute la moins mauvaise formule réside en la combinaison 1) de services publics gratuits couvrant les soins de santé et l'éducation (domaines dans lesquels les besoins individuels sont très différents et doivent être traités de manière différenciée⁸) ainsi qu'une offre culturelle diversifiée ; 2) d'un droit inconditionnel au revenu (différencié en fonction de la tranche d'âge) couvrant les besoins alimentaires et de logement⁹ ; et 3) de compensations particulières pour ceux qui ne peuvent obtenir de revenus supplémentaires par le biais de l'emploi (que ce soit par incapacité ou malchance) ou qui auraient des besoins particuliers aisément identifiables pour lesquels ils ne pourraient être tenus responsables.

UNE APPROCHE PLUS « SUFFISANTISTE » QU'ÉGALITARISTE?

Un reproche qui pourrait être fait à ma conception de l'exploitation est sa focalisation sur les besoins fondamentaux. Suffit-il qu'un individu bénéficie du minimum vital pour qu'il ne soit plus une victime potentielle de l'exploitation? Ne faudrait-il pas plutôt qu'il puisse négocier d'égal à égal avec l'employeur?

Une *garantie* de parfaite égalité de pouvoir de négociation entre employé potentiel et employeur ne pourrait naître que d'une égalité de ressources. Mais dans ce cas, pourquoi l'employé potentiel aurait-il besoin de travailler au service d'un autre? Dans un tel scénario d'égalité des ressources, il n'y aurait sans doute que deux types de travail: le travail indépendant ou le travail coopératif (qui place employés et employeurs sur un parfait plan d'égalité, les deux se confondant dans l'organisation collective). Bien que vraisemblablement souhaitable, ce scénario revêt peu d'intérêt dans l'optique de comprendre et dépasser l'exploitation dans des sociétés largement inégalitaires comme les nôtres. Dans ce contexte non idéal, il importe seulement pour mon propos de mettre en évidence le fait que la garantie d'un seuil de revenu *suffisant* peut entraîner des effets *significatifs* dans la négociation des contrats de travail.

Ainsi, je conçois la lutte contre l'exploitation des travailleurs comme un premier pas dans la poursuite de la justice sociale, même si, comme je l'ai mentionné, cela ne remédierait pas à toutes les injustices. Étant donné le caractère abstrait et particulièrement exigeant de la conception du juste que je serais enclin à défendre (mais ce n'est pas le lieu ici de le faire), il me semble intéressant de voir par moments les choses à rebours: quelles injustices peuvent être « facilement » éliminées? Dans cette perspective, la couverture garantie des besoins fondamentaux constituerait déjà une avancée extrêmement importante, puisqu'elle transformerait profondément les relations de travail en minimisant l'exploitation¹⁰.

Cela ne revient toutefois pas à affirmer que la tâche de l'État devrait se limiter à cette couverture des besoins fondamentaux – « thèse négative » du « suffisantisme » (Casal, 2007). Ce « minimum social » devant empêcher l'exploitation (notamment¹¹) doit plutôt être conçu comme une restriction « suffisantiste » aux inégalités justifiables par la responsabilité personnelle (Casal, 2007 ; Gosseries, 2011). Pour le dire encore autrement, la lutte contre l'exploitation doit être ici comprise comme nécessaire, mais non suffisante du point de vue de la justice sociale. J'ai déjà indiqué qu'un certain égalitarisme des chances pouvait rendre nécessaires des mesures redistributives additionnelles, et je traiterai de l'exploitation dans le cadre plus large de la lutte contre la domination dans la section 7.

REVENU INCONDITIONNEL ET EXPLOITATION

Il existe au moins trois formes que peut revêtir la couverture par l'État des besoins fondamentaux. La première, celle des régimes pleinement socialistes, consiste à abolir la propriété privée et à redistribuer les ressources selon les besoins. La seconde, celle de l'« État social actif » (Vandenbroucke, 2000 ; Vielle, Pochet et Cassiers, 2005), consiste à assurer (outre certains services publics gratuits) un revenu de compensation à ceux qui ne voient pas leurs besoins fondamentaux couverts par les revenus du travail, cela tout en les incitant à (re) trouver un emploi. La troisième, celle de « l'allocation universelle », octroie à tous les citoyens un revenu de base inconditionnel devant permettre de couvrir les besoins fondamentaux non couverts par les services publics gratuits, revenu de base combinable avec les revenus du travail ou certaines compensations supplémentaires pour des désavantages subis.

Je ne discuterai pas ici la première option, qui est la plus radicale et à propos de laquelle l'histoire nous enjoint à une certaine prudence. Rien ne justifie théoriquement que cette piste soit abandonnée, mais il s'agit désormais de mieux prendre en compte, sous un tel régime, la protection des libertés individuelles (y compris les libertés formelles) et l'efficacité économique¹². Ce qui m'intéresse ici, c'est l'idée qu'on puisse remédier à l'exploitation *même en régime capitaliste*¹³, idée qui ne peut être tenue pour évidente tant les marxistes ont toujours insisté sur le lien intrinsèque entre le phénomène de l'exploitation et le régime capitaliste de propriété privée des moyens de production.

Plus accessibles à court terme (pour les sociétés capitalistes, du moins), les deux options « capitalistes » (à entendre dans le sens où ce mot évoque la « propriété privée des moyens de production », et non au sens où il réfère au « libre marché ») doivent être confrontées. La première se focalise sur l'incitation au travail, ce dernier devant permettre aux citoyens de subvenir seuls à l'essentiel de leurs besoins. On peut imaginer un régime de ce type qui serait juste même si un tel régime juste n'existe pour le moment nulle part. La difficulté vient de ce que, sous ce régime, la promotion du travail entre presque systématiquement en conflit avec la garantie des besoins fondamentaux. À moins qu'un revenu minimum très élevé soit légalement imposé – ce qui dissuade les employeurs d'embaucher –, les chômeurs sont peu incités à se mettre à l'emploi (si l'opportunité se présente) dès lors que le premier revenu du travail disponible est souvent à peine supérieur au revenu compensatoire du chômage. De telle sorte que l'espoir de gains d'efficacité (incitation au travail) justifie souvent de manière insidieuse la diminution ou la dégressivité dans le temps de l'allocation de chômage.

Or, l'avantage du régime d'allocation universelle est de faire potentiellement converger la justice et l'efficacité (Van Parijs, 1990), puisque les revenus du travail sont directement additionnés au revenu de base, ce qui, d'un point de vue purement économique, incite davantage au travail (au moins à temps partiel) et permet d'éviter la trappe au chômage induite par la substitution du salaire à l'allocation de chômage sous régime social « actif » (Van Parijs, 1996)¹⁴. Mais cette efficacité économique ne prévaudra qu'à condition que les gains d'opportunités

de loisir offerts par le revenu de base ne sapent pas la motivation (qui n'est pas qu'économique) des citoyens à maintenir la quantité de travail nécessaire au financement de cette allocation universelle, c'est-à-dire à condition qu'elle demeure à un niveau qui la rende soutenable à travers les générations.

Quoi qu'il en soit des mérites respectifs des deux régimes du point de vue de l'efficacité¹⁵ (et donc du caractère *universel*¹⁶ ou non de l'allocation), ce qui importe pour la lutte contre l'exploitation est l'*inconditionnalité* de la couverture des besoins fondamentaux, puisqu'elle offre aux travailleurs la liberté réelle de ne pas travailler, donc un pouvoir de négociation tout autre de leurs conditions salariales.

Cependant, un reproche souvent adressé à l'idée d'un revenu de base inconditionnel (et qui s'applique à toute forme de revenu inconditionnel) consiste à dire qu'il offrirait littéralement aux paresseux le *droit d'exploiter* les travailleurs en vivant, à la manière de parasites, du travail des autres (Elster, 1986, p. 719 ; Van Donselaar, 2009). Il me semble que cette affirmation, défendue de manière sérieuse et élaborée par Gijs Van Donselaar, mobilise une conception peu crédible de l'exploitation. En effet, elle déplace complètement le concept de son origine dans la relation de travail pour faire d'« exploiter » un synonyme de « profiter », alors que les deux ne sont pas équivalents (j'ai déjà rejeté plus haut cette définition « lâche » de l'exploitation). L'exploitation caractérise une situation unilatérale de domination, tandis que le phénomène en question (la liberté réelle de ne pas travailler) est beaucoup plus complexe. On ne peut en effet dire du chômeur qu'il profite du travail d'autrui que dans la mesure où 1) il souhaite effectivement ne pas travailler et 2) il est heureux avec sa maigre allocation, qui lui offre pourtant un désavantage financier comparatif par rapport aux travailleurs (qui n'est compensé par l'avantage relatif lié au temps de loisir que s'il préfère l'oisiveté dans la frugalité). Mais si son chômage est involontaire, par exemple parce qu'il se trouve dans une situation réaliste où le plein-emploi n'est pas réalisé, il subit un désavantage important et ce sont les travailleurs qui profitent du fait qu'ils possèdent cette ressource rare qu'est l'emploi (Van Parijs, 1993 ; 1995) tandis que le chômeur en est privé.

Pour cette raison, et étant donné la difficulté de distinguer le chômage volontaire du chômage involontaire, il ne me paraît pas vraisemblable de caractériser le droit au revenu comme un droit à l'exploitation des travailleurs par les chômeurs¹⁷. Cela reviendrait à se focaliser sur un effet secondaire possible d'une mesure envisagée, effet qu'on gonflerait tant qu'il éclipserait les avantages plus larges visés. Si des injustices de l'ordre du *free riding* pourraient assurément naître de l'implémentation d'un droit au revenu dans un monde non idéal¹⁸, elles me semblent à la fois minimales et largement préférables à l'exploitation engendrée par son absence. Dès lors, si Van Donselaar s'inquiète véritablement de l'exploitation, comme il le prétend, il défend la pire manière imaginable de lutter contre elle : l'obligation de travailler (ou du moins de chercher du travail)¹⁹. Car c'est précisément cette obligation qui offre aux capitalistes le loisir d'exploiter les travailleurs, comme j'ai cherché à le montrer.

MIGRATIONS, CONCURRENCE ET EXPLOITATION

Une autre difficulté surgit cependant, pour ma conception de l'exploitation, du fait que nous vivons dans des sociétés ouvertes, exposées aux flux migratoires, et qu'en aucun lieu du monde les Droits de l'Homme ne coïncident avec ceux du citoyen. En effet, si la couverture des besoins fondamentaux n'est assurée qu'aux citoyens d'un pays, l'exploitation sera reportée sur les non-citoyens. Tant qu'il y aura des individus dans l'obligation matérielle de travailler, il y aura de l'exploitation. En témoignage, parmi d'autres exemples possibles, celui du Qatar, où les citoyens bénéficient tous d'une forme de rente sur l'exploitation nationale des ressources naturelles, ce qui leur offre le pouvoir de ne pas travailler dans des conditions défavorables. En conséquence de quoi ce sont des travailleurs pakistanais, bangladais, népalais et autres qui se tuent (littéralement) à la tâche dans les métiers pénibles ou dangereux qui répugnent (légitimement, étant donné les conditions de travail) aux Qataris.

Par ailleurs, la concurrence entre travailleurs du monde entier dans la lutte des entreprises pour la compétitivité exerce inévitablement une pression à la baisse sur les salaires. Dans le modèle simplifié mettant en scène un seul travailleur face à un seul capitaliste, en admettant que tous deux aient autant besoin de l'autre pour satisfaire certaines préférences non fondamentales, l'équilibre de la négociation débouche sur une situation d'avantages mutuels avec un salaire équitable pour le travailleur. Cependant, ce genre de situation mutuellement avantageuse n'est pas toujours équitable (Barry, 1989 ; 1995). Plus le pouvoir de négociation du travailleur diminue, plus la relation de travail tend à l'exploitation, même s'il peut demeurer dans l'avantage des deux parties de signer un contrat²⁰. Or, plus les travailleurs seront mis en concurrence avec des travailleurs étrangers possédant un pouvoir de négociation presque nul, plus ils verront leur propre pouvoir de négociation s'amoindrir, avec les conséquences que l'on connaît pour l'exploitation.

Cela ne signifie évidemment pas qu'il faille renoncer à lutter contre l'exploitation des travailleurs nationaux, mais il convient de toujours situer ce combat dans un cadre plus large, qui est celui de la justice mondiale. C'est pourquoi, plus que jamais, les aspirations marxistes (d'inspiration du moins) à davantage de justice sociale doivent se situer dans une perspective internationale, et les partis politiques de cette obédience, s'unir pour présenter des projets supranationaux communs (à l'échelle européenne, par exemple, puis globale).

Mais le problème ne naît pas uniquement de la mondialisation économique. Même au sein d'un hypothétique contexte national politiquement homogène (où tous les travailleurs possèderaient les mêmes droits), si certains emplois sont à la fois rares, pénibles et pourtant enviés (c'est imaginable si la plupart des gens préfèrent le travail, quel qu'il soit, à l'oisiveté), la concurrence entre travailleurs potentiels entraînerait également une pression à la baisse sur les salaires. Dans cette situation, un employé obtiendrait certainement, pour ce type de travail, un salaire inférieur à ce qu'il obtiendrait dans une situation de non-concurrence.

Cela revient-il à dire qu'il pourrait être exploité bien que la satisfaction de ses besoins fondamentaux soit assurée par l'État? Je ne pense pas, car il posséderait toujours la liberté réelle de refuser l'emploi et d'attendre par exemple des circonstances plus favorables pour travailler.

Si l'on retourne dans un monde ouvert, avec économie globale, cette liberté est en danger, puisque la recherche nationale de compétitivité risque de remettre en question l'inconditionnalité du droit au revenu. Mais ce danger guette toute proposition visant un progrès social en contexte mondialisé. Il est évident que pour gagner en crédibilité économique (mais pas en faisabilité politique), toute proposition de ce type doit s'appliquer au niveau de pouvoir politique le plus élevé possible. Appliqué au niveau mondial, le droit inconditionnel à un revenu de subsistance, bien plus fondamental que le Droit de l'Homme à un travail (Article 23 de la Déclaration universelle), serait un formidable remède à l'exploitation. Au niveau national, il sera toujours menacé, à l'instar de toute mesure redistributive visant la justice sociale.

Sur cette base, on peut distinguer de ma première définition (« basique ») de l'exploitation une définition « forte » : *une personne est exploitée si le salaire qu'elle obtient en échange de son travail est inférieur à ce qu'il serait si elle – et ses pairs, en tous lieux – bénéficiaient d'une garantie inconditionnelle de voir leurs besoins fondamentaux couverts (par l'État). (Et si elle ne dispose, par ailleurs, ni de revenus liés au capital, ni d'un talent rare.)*

Je ne suis pas sûr que cette définition invalide la précédente (raison pour laquelle je ne la substitue pas à l'autre), mais elle permet mieux d'identifier les deux mécanismes fondamentaux à l'origine de l'exploitation : l'obligation matérielle de travailler et la concurrence entre travailleurs soumis à différents régimes de droits – ce qui plaide pour l'instauration d'un Droit de l'Homme à un revenu de subsistance.

EXPLOITATION ET DOMINATION

Partant de la définition de l'exploitation que j'ai défendue ici, la couverture inconditionnelle des besoins fondamentaux apparaît comme une mesure contribuant à réduire la vulnérabilité²¹ des travailleurs et demandeurs d'emploi, celle-là même qui leur fait accepter des contrats injustes dans un monde caractérisé par les inégalités de pouvoir. Ce programme pourrait assurément être situé dans un projet plus général de lutte contre la domination²². Que gagne-t-on dès lors à se focaliser sur l'exploitation dans les relations de travail, alors que cette même couverture des besoins fondamentaux pourrait également réduire la domination familiale, par exemple?

La raison pour laquelle je me suis personnellement intéressé à cette question précise, c'est qu'il me semble que nous vivons à une époque qui place travailleurs et chômeurs sous une double pression particulièrement forte. D'une part, celle inhérente à un monde du travail extrêmement compétitif, qui est la source de nombreuses souffrances (Dejours, 1998). D'autre part, la pression morale

culpabilisante que nos sociétés exercent sur les chômeurs, qu'on a tôt fait d'accuser, sans trop se mettre à leur place, de profiter du système. Or, outre le fait que ces jugements sont déplacés dans un contexte d'absence de plein-emploi, étant donné la difficulté de distinguer le chômage volontaire de l'involontaire, les jugements moraux que l'on peut se permettre à l'égard de ceux qui refusent le monde du travail dépendent pour large part du type d'emplois qui s'offrent à eux. D'où la nécessité, à mes yeux, de repenser le concept d'exploitation afin de réfléchir à ce que l'on peut attendre des citoyens face à l'emploi. La mise en évidence de l'origine de l'exploitation dans l'obligation *matérielle* de travailler même, me semble-t-il, à considérer avec prudence l'obligation *morale* de travailler. Ce qui précède souligne donc toute l'importance de l'*inconditionnalité* des allocations sociales, qu'elles soient universelles ou adressées aux seuls chômeurs.

Je ne prétends cependant pas qu'un tel droit inconditionnel au revenu soit le remède à tous les maux des travailleurs. Comme je l'ai indiqué, la lutte contre l'exploitation possède une dimension « suffisantiste », tandis que la justice sociale réclame sans doute davantage encore (Casal, 2007). En régime capitaliste, ceux qui sont privés de capital seront fréquemment (mais pas forcément dans un régime capitaliste juste) plus vulnérables que les capitalistes. Dès lors, une meilleure redistribution du capital (que permettrait d'ailleurs un revenu de base financé par la taxation des dons et héritages) s'avère également nécessaire à l'égalité des chances. Les syndicats doivent également continuer à jouer leur rôle important consistant à réduire les conséquences de la concurrence internationale entre travailleurs potentiels et augmenter le pouvoir de négociation collectif de ces derniers²³.

Tout mécanisme permettant de réduire la domination me paraît cohérent avec la lutte contre l'exploitation des travailleurs sur laquelle s'est focalisé cet article. J'espère néanmoins avoir atteint, au moyen de cette focalisation, deux objectifs : 1) offrir une meilleure conceptualisation de la notion de sens commun d'exploitation ; 2) donner à réfléchir aux conditions auxquelles il peut être moralement adéquat d'« activer » les citoyens, selon la formule à la mode, c'est-à-dire de les contraindre à trouver un emploi.

CONCLUSION

Il me paraît encore utile de penser l'exploitation dans la relation de travail, mais sans se référer à la propriété de soi et des produits de son travail. C'est l'obligation de travailler qui est à la source de l'exploitation. Quant au critère que j'ai proposé pour déterminer s'il y a ou non exploitation (la rémunération qu'accepterait le travailleur s'il disposait de la liberté réelle de refuser l'emploi), il demeure en partie indéterminé, car subjectif, ce qui est cohérent avec la mise en avant par Marx de la dimension fondamentalement subjective du travail (Henry, 2008, p. 30), et non problématique si les salaires sont déterminés par un marché et non directement par l'État.

Lutter contre l'exploitation constituerait un premier pas fondamental dans la lutte pour la justice sociale. Il resterait cependant à compenser les inégalités d'avantages liées aux circonstances subies. Ce n'est pas parce qu'un travailleur n'est pas exploité que sa rémunération est juste. Il n'est pas juste qu'il soit avantage eu égard à ses talents naturels (qui lui permettent de monnayer ses forces productives à plus haut prix), étant donné que personne ne devrait être désavantagé, dans une société juste, pour de telles raisons moralement arbitraires. Par ailleurs, bien que la couverture inconditionnelle des besoins fondamentaux entraînerait certainement un plus grand partage du travail (l'obligation de travailler disparaissant), s'il devait subsister un chômage involontaire, celui-ci devrait également être compensé par une redistribution, dès lors que les travailleurs tireraient un avantage d'un bien rare et envié (l'emploi) dont d'autres se voient privés contre leur gré.

Je rappelle enfin que la définition de l'exploitation que j'ai proposée n'apparaît plus comme inhérente au capitalisme (entendu strictement comme régime de propriété privée des moyens de production). Sa disparition peut tout aussi bien advenir en régime capitaliste qu'en régime socialiste de marché (propriété publique des moyens de production et allocation des emplois, notamment, par un marché), les deux possédant leurs avantages et défauts respectifs²⁴ (notamment la réticence individuelle à la taxation en régime capitaliste et la concentration du pouvoir en régime socialiste). Peut-être qu'un régime socialiste subirait moins de résistances individuelles à la couverture des besoins fondamentaux de tous. Mais peut-être aussi qu'une société capitaliste juste est plus accessible aux pays capitalistes que l'idéal socialiste, dès lors que ce dernier impliquerait une douloureuse « confiscation » de la propriété (ou nationalisation) des moyens de production.

Quoi qu'il en soit, l'idéal communiste d'une société ayant pour emblème « À chacun selon ses besoins » pourrait prendre la forme suivante (que la propriété des moyens de production soit privée ou collective) : à chacun selon ses besoins via l'État, et à chacun selon ses préférences sur le marché²⁵ (de l'emploi et des biens).

NOTES

- ¹ Cette formulation de l'argument, due à Gerald Allan Cohen, n'est pas celle de Marx qui, comme le montre habilement Cohen, est encombrée par la théorie de la valeur travail (dont je ne parlerai pas ici). Mais cette formulation-ci est celle qui permet d'exprimer le plus clairement le problème *normatif* de l'exploitation. Qu'elle soit ou non fidèle à Marx, c'est elle qui exprime sans doute le mieux la condamnation marxiste du capitalisme.
- ² Du moins son pouvoir d'attraction est contestable, car il n'est peut-être pas possible de réfuter un tel principe posé a priori par certains. Voir Cohen, 1995, pp. 229-230. L'ensemble du paragraphe qui suit est une libre reformulation des arguments de Cohen, 1995, pp. 229-244.
- ³ Au contraire de ce que prétend Nozick. Mais Nozick transforme subrepticement la maxime kantienne en une maxime libertarienne lorsqu'il la reformule en disant qu'on ne peut traiter une personne comme un moyen *sans son consentement* (ce qui est très différent et relève de questions liées à la légitimité de l'autorité politique en général). Voir Nozick, 1974, p. 31 et Cohen, 1995, pp. 241-243. Locke, lui, avait résolu la question des taxes en assimilant le consentement individuel au consentement de la majorité, ce qui n'est pas évident, et que refuse Nozick. Voir Locke, 1690 ; Barry, 1996, p. 28.
- ⁴ Un égalitariste libéral comme Brian Barry (1996, p. 28), pour sa part, estime n'avoir nullement besoin de ces arguments. L'hypothèse de propriété de soi lui paraît une aberration, le résultat d'une « confusion linguistique » qu'il faut « être fou pour croire ». Mais comme l'explique Cohen (1996) en réponse à cette attaque, le principe de propriété de soi et des fruits de son travail, aussi absurde qu'il puisse paraître à Barry, possède une force d'attraction énorme dans le débat public et n'a pas fini de séduire des esprits réfractaires à l'idée de redistribution (particulièrement aux États-Unis, ajouterais-je). En outre, s'il est évident que tous les marxistes sérieux sont favorables à un certain impôt sur le travail, il est vrai que les principes normatifs qui les guident, en particulier celui qui concerne l'exploitation, manquent de clarté. C'est pourquoi il faut prendre au sérieux le « défi libertarien » et refonder l'égalitarisme marxiste.
- ⁵ Richard Arneson (1981, p. 212), notamment, dit de l'appropriation de la plus-value qu'elle n'est injuste qu'à condition 1) qu'elle soit le résultat d'une distribution inégale du « pouvoir social » et 2) qu'elle entraîne une distribution extrêmement inégalitaire des avantages économiques.
- ⁶ Contre l'idée répandue selon laquelle les questions distributives importaient peu à Marx, Arneson (1981, pp. 222-225) défend une compréhension de l'exploitation telle que conçue par Marx comme ancrée dans des préoccupations distributives.
- ⁷ Formulation légèrement modifiée par rapport à celle de Cohen.
- ⁸ En ce qui concerne l'enseignement, j'entends ce traitement différencié non dans le sens où certaines personnes n'ont pas besoin d'éducation, mais dans le sens où l'école publique est censée traiter les élèves selon leurs besoins particuliers, afin de tous les mener (au moins) vers un certain seuil commun.
- ⁹ On pourrait également poser qu'il est préférable de garantir ces besoins par une offre publique de logements sociaux et des cantines publiques, mais je ne suis pas sûr que les gains attendus d'une telle intrusion de l'État dans les choix individuels justifient de sacrifier une telle dose de liberté individuelle.
- ¹⁰ Bien entendu, ce n'est pas la seule manière de lutter contre l'exploitation. Une option moins coûteuse pour l'État consiste à imposer aux entreprises un salaire minimum suffisamment élevé. Mais ce mécanisme est insensible au coût subjectif du travail (dimension importante de l'exploitation) et entre en tension, étant donné l'augmentation du coût du travail, avec une politique de plein-emploi en régime capitaliste, ce qui risque d'augmenter les injustices entre travailleurs et chômeurs involontaires.
- ¹¹ Je ne considère pas ici d'autres effets bénéfiques pour la justice de la couverture garantie des besoins fondamentaux, comme le respect de soi ou la valeur des libertés politiques.
- ¹² Des propositions très stimulantes allant dans ce sens ont été proposées par Joseph Carens (1981) et John Roemer (1994). Elles sont évaluées de manière critique dans Cohen (2009).

- ¹³ Ce qui ne revient pas à affirmer que je trouve cette option préférable dans l'idéal.
- ¹⁴ L'autre avantage comparatif étant l'effet positif escompté sur le partage du travail.
- ¹⁵ Une discussion intéressante des mérites respectifs du point de vue de l'efficacité de plusieurs modèles théoriques (revenu de base, subventions salariales et régimes mixtes) satisfaisant la même exigence de justice (maximin d'opportunités de revenu et de loisir) est proposée par Van der Veen, 2004. Il suggère que le choix entre différents modèles possibles est davantage une question de choix politiques que de principes de justice (et fournit des éléments pour nourrir la réflexion publique). Outre la justice et l'efficacité, il me semble que d'autres critères pourraient entrer en considération dans le choix d'une formule plutôt que de l'autre, comme une réflexion collective sur le rapport au travail qu'une société entend entretenir.
- ¹⁶ À entendre dans le sens d'une allocation « distribuée à tous, au sein de la communauté politique en question », comme le suffrage universel.
- ¹⁷ Van Parijs (1993) juge même plutôt que ce sont les travailleurs qui exploitent les chômeurs en tirant profit de cette ressource rare qu'est l'emploi. Mais j'ai déjà mentionné que je ne souscrivais pas à cette compréhension élargie de l'exploitation.
- ¹⁸ C'est pourquoi il y aurait sans doute lieu de promouvoir également un certain *éthos* de justice qui inciterait à travailler ceux qui en sont capables et ont l'opportunité d'apporter leur contribution à la coopération sociale dans un emploi ne relevant pas de l'exploitation.
- ¹⁹ Je ne prétends pas que Van Donselaar défend l'obligation de travailler en elle-même, mais il me semble qu'il s'agit de la conséquence logique de son refus de l'inconditionnalité. Il affirme que les revenus du chômage devraient être tels que personne ne puisse faire usage de leur garantie comme incitation à refuser un emploi (Van Donselaar, 2009, p. 149). C'est rendre les travailleurs potentiels encore plus vulnérables à l'exploitation.
- ²⁰ Barry (1995, pp. 39-46) évoque, pour illustrer la différence entre justice et avantage mutuel, l'exemple d'Amérindiens ayant accepté de signer des traités iniques avec les colons parce que cela les exposait à des conséquences moins graves que s'ils ne signaient pas, bien qu'il y eût injustice manifeste.
- ²¹ Cette notion, qu'il situe au cœur du phénomène d'exploitation, est au centre de la défense des droits sociaux par Robert Goodin (1988). Je remercie le lecteur anonyme qui a attiré mon attention sur ce point. Plus récemment, Nicholas Vrousalis (2013) a développé une conception de l'exploitation centrée sur cette même notion de vulnérabilité et sur les rapports de domination entre capitalistes et travailleurs. Il estime pour sa part que l'exploitation ne peut être combattue sous régime capitaliste, car il subsistera toujours une forme de domination des capitalistes sur les travailleurs, rendant ces derniers vulnérables à l'exploitation.
- ²² Voir Pettit, 1997. (Ici aussi je remercie le lecteur anonyme.) Pettit (2007) suggère d'ailleurs une affinité de visées entre l'idée d'un revenu de base inconditionnel et celle du républicanisme comme non-domination.
- ²³ On peut cependant s'attendre à ce que, dans un tel cas de figure (droit au revenu de subsistance), l'importance des syndicats pour les travailleurs diminue. C'est une des (mauvaises) raisons pour lesquelles la plupart des syndicats sont opposés au revenu de base. Voir Vanderborght, 2006.
- ²⁴ Pour peu que l'on compare deux modèles optimaux du point de vue de la justice sociale. Beaucoup trop souvent, on fait l'erreur de comparer un modèle socialiste optimal avec la réalité capitaliste truffée d'injustices, ou inversement le modèle idéal de libre marché en concurrence parfaite avec la planification soviétique historique, par exemple. Voir Van Parijs, 1995.
- ²⁵ Il n'est pas nécessaire d'être naïf par rapport aux mobiles (cupidité et peur, principalement) que mobilise la logique de marché pour la préférer, pour des raisons de justice (la liberté d'occupation, notamment), d'efficacité et de plausibilité à l'économie planifiée. Voir Cohen, 1995, pp. 245-265 ; 2009.

BIBLIOGRAPHIE

Arneson, Richard, "What's Wrong with Exploitation?," *Ethics*, vol. 91, no. 2, 1981, pp. 202-227.

Arneson, Richard, "Equality and Equal Opportunity for Welfare," *Philosophical Studies*, vol. 56, no. 1, 1989, pp. 77-93.

Barry, Brian, *Theories of Justice. A Treatise on Social Justice. Vol. I*, Londres, Hemel Hempstead, Harvester Wheatsheaf, 1989.

Barry, Brian, *Justice as Impartiality. A Treatise on Social Justice. Vol. II*, Oxford, Oxford University Press, 1995.

Barry, Brian, "You Have to Be Crazy to Believe It," *Times Literary Supplement*, 25 octobre 1996, pp. 28.

Carens, Joseph, *Equality, Moral Incentives and the Market*, Chicago, University of Chicago Press, 1981.

Casal, Paula, "Why Sufficiency Is Not Enough," *Ethics*, vol. 117, no. 2, 2007, pp. 296-326.

Cohen, Gerald A., *History, Labour and Freedom. Themes from Marx*, Oxford, Oxford University Press, 1988.

Cohen, Gerald A., "On the Currency of Egalitarian Justice," *Ethics*, vol. 99, no. 4, 1989, pp. 906-944.

Cohen, Gerald A., *Self-ownership, Freedom and Equality*, Cambridge, Cambridge University Press, 1995.

Cohen, Gerald A., "Self-ownership and the Libertarian Challenge," *Times Literary Supplement*, 8 novembre 1996.

Cohen, Gerald A., *Pourquoi pas le socialisme ?*, Paris, L'Herne, [2009] 2010.

Dejours, Christophe, *Souffrance en France. La banalisation de l'injustice sociale*, Paris, Seuil, 1998.

Dworkin, Ronald, "What is Equality? Part 1: Equality of Welfare," *Philosophy & Public Affairs*, vol. 10, no. 3, 1981a, pp. 185-246.

Dworkin, Ronald, "What is Equality? Part 2: Equality of Resources," *Philosophy & Public Affairs*, vol. 10, no. 4, 1981b, pp. 283-345.

Elster, Jon, "Exploring Exploitation," *Journal of Peace Research*, vol. 15, no. 1, 1978, pp. 3-17.

Elster, Jon, "Comment on Van der Veen and Van Parijs," *Theory and Society*, vol. 15, no. 5, 1986, pp. 709-721.

Goodin, Robert, *Reasons for Welfare: The Political Theory of the Welfare State*, Princeton, Princeton University Press, 1988.

Gosseries, Axel, « Qu'est-ce que le suffisantisme ? », *Philosophiques*, vol. 38, no. 2, 2011, pp. 465-491.

Henry, Michel, *Le socialisme selon Marx*, Cabris, Sulliver, 2008.

Kymlicka, Will, *Contemporary Political Philosophy: An Introduction*, Oxford, Oxford University Press, 2002.

Locke, John, *Le second traité du gouvernement. Essai sur la véritable origine, l'étendue et la fin du gouvernement civil*, Paris, Presses universitaires de France, [1690] 1994.

Marx, Karl, *Le Capital. Livre 1* (2 vol.), Paris, Flammarion, [1867] 1985.

Nozick, Robert, *Anarchie, État et utopie*, Paris, Presses universitaires de France, [1974] 2003.

Pettit, Philip, *Republicanism: A Theory of Freedom and Government*, Oxford, Oxford University Press, 1997.

Pettit, Philip, "A Republican Right to Basic Income?," *Basic Income Studies*, vol. 2, no. 2, 2007.

Rawls, John, *Théorie de la Justice*, Paris, Gallimard, [1971] 1991.

Roemer, John, *A General Theory of Exploitation and Class*, Cambridge, Harvard University Press, 1982.

Roemer, John, *Egalitarian Perspectives: Essays in Philosophical Economics*, Cambridge, Cambridge University Press, 1994.

Roemer, John, *Theories of Distributive Justice*, Cambridge, Harvard University Press, 1996.

Roemer, John, *Equality of Opportunity*, Cambridge, Harvard University Press, 1998.

Sen, Amartya, *Inequality Reexamined*, Cambridge, Harvard University Press, [1992] 1995.

Sen, Amartya, *L'idée de justice*, Paris, Flammarion, [2009] 2010.

Spitz, Jean-Fabien, *Abolir le hasard? Responsabilité individuelle et justice sociale*, Paris, J. Vrin, 2008.

Vandenbroucke, Frank, *Op zoek naar een redelijke utopie. De actieve welvaartstaat in perspectief*, Antwerpen, Garant, 2000.

Vanderborght, Yannick, "Why Trade Unions Oppose Basic Income," *Basic Income Studies*, vol. 1, no. 1, 2006.

Van der Veen, Robert, "Basic Income Versus Wage Subsidies: Competing Instruments in an Optimal Tax Model with a Maximin Objective," *Economics and Philosophy*, no. 1, 2004 pp. 147-183.

Van Donselaar, Gijs, *The Right to Exploit: Parasitism, Scarcity, and Basic Income*, Oxford, Oxford University Press, 2009.

Van Parijs, Philippe, "The Second Marriage of Justice and Efficiency," *Journal of Social Policy*, vol. 19, no. 1, 1990, pp. 1-25.

Van Parijs, Philippe, *Marxism Recycled*, Paris/Cambridge, Maison des sciences de l'homme/Cambridge University Press, 1993.

Van Parijs, Philippe, *Real Freedom for All: What (if Anything) Can Justify Capitalism?*, Oxford, Oxford University Press, 1995.

Van Parijs, Philippe, « De la trappe au socle. L'allocation universelle contre le chômage », *Revue suisse de science politique*, Genève, vol. 2, no. 1, 1996, pp. 103-117.

Vielle, Pascale, Pochet, Philippe et Isabelle Cassiers, *L'État social actif : vers un changement de paradigme ?*, Bruxelles, Lang, 2005.

Vrousalis, Nicholas, "Exploitation, Vulnerability, and Social Domination," *Philosophy and Public Affairs*, vol. 41, no. 2, 2013, pp. 131-157.

DOSSIER

ATTITUDES, VALEURS ET ENVIRONNEMENT

ANTOINE C. DUSSAULT

DOCTORANT EN PHILOSOPHIE, UNIVERSITÉ DE MONTRÉAL

L'éthique de l'environnement, telle qu'elle s'est développée en Amérique du Nord depuis la fin des années 1970, a toujours accordé une grande place aux questions concernant la nature et les fondements des concepts et des jugements de valeur, des questions traditionnellement associées à l'éthique fondamentale. Bien que cette discipline porte sur un enjeu d'éthique appliquée, soit celui de caractériser le type de relation que doivent entretenir les agents moraux avec leur environnement naturel, un de ses projets centraux—celui d'expliquer et défendre l'idée selon laquelle la nature a une valeur finale et de montrer que la reconnaissance de cette valeur a des conséquences éthiques importantes—soulève plusieurs questions qui dépassent le cadre habituel de l'éthique appliquée. Ces questions concernent la notion de valeur finale et le rôle qu'elle occupe en éthique, des enjeux qui ont intéressé autant des spécialistes d'éthique fondamentale que d'éthique de l'environnement.

L'une de ces questions est celle de la relation entre les concepts de valeur finale et de valeur intrinsèque, c'est-à-dire celle de savoir si une chose peut avoir de la valeur pour elle-même (*for its own sake*) en vertu de ses relations, ou si elle ne peut avoir une telle valeur qu'en vertu de ses propriétés intrinsèques (ex. : Bradley, 2001 ; Green, 1996 ; Elliot, 1997 ; 1992 ; O'Neill, 1992 ; Thompson, 1990 ; Gunn, 1980). La réponse donnée à cette question a des conséquences directes sur la possibilité d'accorder une valeur finale à certains environnements en vertu de leur caractère sauvage, ou à certaines espèces en vertu de leur rareté. Une autre question concerne le débat entre monistes et pluralistes à propos des valeurs, et consiste à déterminer si toutes les valeurs se réduisent ultimement à une seule ou s'il existe plusieurs catégories de valeurs irréductibles les unes aux autres (ex. : Norton, 1995 ; Callicott, 1994 ; 1990 ; Varner, 1991 ; Stone, 1988 ; 1987). Cette problématique touche notamment l'enjeu de savoir jusqu'à quel point les considérations relevant de l'éthique des relations interhumaines, et celles relevant de

l'éthique animale et environnementale peuvent et doivent être unifiées, ou si elles reposent sur des bases radicalement distinctes. Une autre question est celle du statut ontologique des valeurs et de la relation entre les propriétés de valeur et les états mentaux (préférences, émotions, attitudes) des évaluateurs et évaluatrices. Cette question est liée à l'enjeu de l'anthropocentrisme et à la question de savoir jusqu'à quel point et en quel sens les valeurs autour desquelles s'articule l'éthique de l'environnement peuvent et doivent être indépendantes des sensibilités humaines (ex. : McShane, 2011; Frierson, 2010; 2007; Elliot, 1996; 1992; Callicott, 1995; 1992a; 1992b; Rolston, 1994; 1981; Miller, 1982). Finalement, une des questions d'éthique fondamentale qui a suscité le plus d'intérêt en éthique de l'environnement concerne la nature des notions de tort et de bénéfice et le type d'entités auxquelles ces notions peuvent s'appliquer (ex. : Varner, 1998; 1990; Salthe et Salthe, 1989; Cahen, 1988; Taylor, 1986; 1981; Regan, 1983; Feinberg, 1974). Les réponses données à cette question ont des conséquences directes sur la possibilité de reconnaître une considérabilité morale (Goodpaster, 1993; 1979; 1978) aux animaux sensibles, aux êtres vivants non conscients et aux tous écologiques comme les communautés et les écosystèmes.

Bien que les spécialistes d'éthique de l'environnement et d'éthique fondamentale démontrent un intérêt commun pour ces questions touchant la nature de la valeur, il est surprenant de constater la rareté de leurs échanges à ce sujet. Les deux champs de recherche ont encore aujourd'hui tendance à mener leurs investigations isolément, chacun se concentrant sur son propre réseau d'auteurs, tirant peu profit des contributions issues de l'autre domaine. C'est de ce constat qu'est née l'idée d'organiser l'atelier « Attitudes, valeurs et environnement », tenu à l'Université de Montréal le 21 février 2014, et dont sont dérivés les articles présentés dans ce dossier. L'objectif de l'atelier était de discuter de questions se situant à l'intersection de l'éthique fondamentale et de l'éthique de l'environnement en combinant les perspectives de ces deux champs de recherche, ce que reflètent les présentes contributions.

Les deux premiers articles portent sur la relation entre les préférences et les valeurs. "Thinking Globally, Acting Locally: Partiality, Preferences and Perspective," par Graham Oddie, discute des manières par lesquelles l'analyse de la valeur en termes d'attitudes appropriées (*fitting-attitude analysis of value*), la AAA pour être plus concis, en s'associant à la théorie de la valeur conçue comme la richesse (définie plus ou moins comme la variété unifiée) (Kelly, 2014; 2003; Mikkelson, 2014; 2011a; Bradley, 2001; Miller, 1982), peut accommoder l'intuition selon laquelle un certain degré de partialité préférentielle est légitime chez les évaluatrices et évaluateurs. La partialité préférentielle semble être en tension avec la AAA, puisque celle-ci paraît requérir une isomorphie complète entre les réponses évaluatives des agents. Oddie développe trois manières par lesquelles cette exigence pourrait être assouplie et précise dans quelle mesure celles-ci peuvent être complémentaires. La deuxième contribution, « Sur la symétrie présumée entre valeurs et préférences », par Mauro Rossi, aborde le rapport entre valeurs et préférences dans l'analyse de la relation de parité, une relation comparative non standard où aucun des deux éléments comparés n'est

supérieur à l'autre en valeur, et où leur valeur n'est pas pour autant égale. Adoptant lui aussi le cadre de la AAA, Rossi défend l'interprétation de la parité en termes de permissibilité proposée par Rabinowicz (2008), contre la critique qui lui a été adressée par Gustafsson (2013) sur la base d'une symétrie présumée entre valeurs et préférences.

Les deux articles suivants portent spécifiquement sur la théorie de la valeur conçue comme la richesse. "Richness Theory: From Value to Action," par Gregory M. Mikkelsen, poursuit les travaux entrepris par l'auteur sur cette théorie (Mikkelsen et Champan, à paraître; Mikkelsen, à paraître; 2011a; 2011b), et en développe le volet pratique. Mikkelsen compare la richesse générée par deux futurs possibles issus des choix que feront les humains dans les prochaines années, et montre que le scénario hypothétique où les humains diminuent leur empreinte écologique afin de n'utiliser chaque année que les ressources produites annuellement par la terre génère plus de richesse que le scénario *business as usual*, où les humains ne changent rien à leurs manières de vivre. S'appuyant ensuite sur la théorie conséquentialiste de la décision élaborée par Jackson (1991), Mikkelsen fait valoir que les humains ont le devoir de mettre en œuvre les moyens de réaliser le premier scénario, même si ce projet a moins de chances de se réaliser que le scénario *business as usual*. Le quatrième article, "Value Monism, Richness, and Environmental Ethics," par Christopher Kelly, défend une théorie *moniste* de la valeur comme la richesse, et en montre les implications pour l'éthique de l'environnement. Kelly met d'abord en évidence plusieurs limites du pluralisme en éthique, et défend l'idée selon laquelle une théorie moniste capable d'unifier les diverses intuitions des évaluateurs et évaluatrices serait préférable à une théorie pluraliste. Puisant dans les travaux esthétiques de plusieurs penseurs classiques comme Leibniz, Hutcheson et Aristote, et ceux plus récents de Nozick (1981) et Osborne (1953), Kelly élabore ensuite une définition de la richesse comme la variété unifiée et illustre son caractère intuitif par des exemples biologiques, esthétiques et relevant de la méthodologie scientifique. Kelly montre finalement comment la théorie de la valeur comme la richesse explique, mieux selon lui que les théories anthropocentristes et les principales théories d'éthique animale, les intuitions les plus fortes concernant la valeur finale des animaux, des organismes vivants, des écosystèmes et de la *wilderness*.

La cinquième et dernière contribution, "Individualist Biocentrism vs. Holism Revisited," par Katie McShane, reconsidère le débat entre biocentristes et écocentristes sur la possibilité d'attribuer des intérêts aux tous écologiques comme les écosystèmes et de leur reconnaître, sur cette base, une considérabilité morale (Agar, 2001; Varner, 1998; Salthe et Salthe, 1989; Cahen, 1988; Taylor, 1986). Après avoir mis en évidence le fait que le contexte des changements climatiques réinscrit le débat individualisme/holisme à l'ordre du jour, McShane montre que certaines avancées en philosophie de la biologie, en médecine et en biologie évolutionnaire affaiblissent la distinction entre individus et tous sur laquelle reposent les principaux arguments formulés par les biocentristes contre l'holisme en éthique de l'environnement.

En complément à ce dossier, le numéro de l'été 2014 de la revue *Pensées canadiennes* a publié des entrevues réalisées parallèlement à l'atelier « Attitudes, valeurs et environnement » et portant sur des thèmes connexes. Grâce à l'initiative de Vanessa Deverell, Maude Ouellette-Dubé et Laurent Vachon Roy, éditrices et éditeur en chef de cette revue, on peut lire dans ce numéro les propos (en anglais et en français) de Graham Oddie, Gregory Mikkelson, Christopher Kelly et Katie McShane (<http://www.penseescanadiennes.com/>).

BIBLIOGRAPHIE

Agar, Nicholas, *Life's Intrinsic Value Science, Ethics, and Nature*, New York, Columbia University Press, 2001.

Bradley, Ben, "The Value of Endangered Species," *Journal of Value Inquiry*, vol. 35, no. 1, 2001, pp. 43-58.

Cahen, Harley, "Against the Moral Considerability of Ecosystems," *Environmental Ethics*, vol. 10, no. 3, 1988, pp. 195-216.

Callicott, J. Baird, "The Case Against Moral Pluralism," *Environmental Ethics*, vol. 12, no. 2, 1990, pp. 99-124.

Callicott, J. Baird, "Rolston on Intrinsic Value: A Deconstruction," *Environmental Ethics*, vol. 14, no. 2, 1992a, pp. 129-143.

Callicott, J. Baird, "Can a Theory of Moral Sentiments Support a Genuinely Normative Environmental Ethic?," *Inquiry*, vol. 35, no. 2, 1992b, pp. 183-198.

Callicott, J. Baird, "Moral Monism in Environmental Ethics Defended," *Journal of Philosophical Research*, no. 19, 1994, pp. 51-60.

Callicott, J. Baird, "Intrinsic Value in Nature: A Metaethical Analysis," *Electronic Journal of Analytic Philosophy*, vol. 3, no. 5, 1995.

Elliot, Robert, "Intrinsic Value, Environmental Obligation and Naturalness," *The Monist*, vol. 75, no. 2, 1992, pp. 138-160.

Elliot, Robert, "Facts About Natural Values," *Environmental Values*, vol. 5, no. 3, 1996, pp. 221-234.

Elliot, Robert, *Faking Nature: The Ethics of Environmental Restoration*, London/New York, Routledge, 1997.

Feinberg, Joel, "The Rights of Animals and Unborn Generations," in *Philosophy and Environmental Crisis*, Athens, University of Georgia Press, 1974, pp. 43-68.

Frierson, Patrick R., "Metastandards in the Ethics of Adam Smith and Aldo Leopold," *Environmental Ethics*, vol. 29, no. 2, 2007, pp. 171-191.

Frierson, Patrick R., "Smithian Intrinsic Value," in Brown, Vivienne et Samuel Fleischacker (eds.), *The Philosophy of Adam Smith: Essays Commemorating the 250th Anniversary of the Theory of Moral Sentiments*, Adam Smith Review 5, London/New York, Routledge, 2010.

Goodpaster, Kenneth E., "On Being Morally Considerable," *The Journal of Philosophy*, vol. 75, no. 6, 1978, pp. 308-325.

Goodpaster, Kenneth E., "From Egoism to Environmentalism," in Goodpaster, Kenneth E. et Sayre Kenneth M. (eds.), *Ethics and Problems of the 21st Century*, Notre Dame, University of Notre Dame Press, 1979, pp. 21-35.

Goodpaster, Kenneth E., "Moral Consideration and the Environment: Perception, Analysis, and Synthesis," *Topoi*, vol. 12, no. 1, 1993, pp. 5-20.

Green, Karen, "Two Distinctions in Environmental Goodness," *Environmental Values*, vol. 5, no. 1, 1996, pp. 31-46.

Gunn, Alastair S., "Why Should We Care About Rare Species?," *Environmental Ethics*, vol. 2, no. 1, 1980, pp. 17-37.

Gustafsson, Johan E., "Value-Preference Symmetry and Fitting-Attitude Accounts of Value Relations," *Philosophical Quarterly*, vol. 63, no. 252, 2013, pp. 476-491.

Jackson, Frank, "Decision-theoretic Consequentialism and the Nearest and Dearest Objection," *Ethics*, vol. 101, no. 3, 1991, pp. 461-482.

Kelly, Chris, "Value Monism, Richness and Environmental Ethics," *Les ateliers de l'éthique/The Ethics Forum*, vol. 9, no. 2, 2014, pp. 110-129.

Kelly, Crhistopher, *A Theory of the Good*, Ph.D. Thesis, University of Colorado, Boulder, CO, 2003.

McShane, Katie, "Neosentimentalism and Environmental Ethics," *Environmental Ethics*, vol. 33, no. 1, 2011, pp. 5-23.

Mikkelson, Gregory M., "Richness Theory: From Value to Action", *Les Ateliers de l'éthique/The Ethics Forum*, vol. 9, no. 2, 2014, pp. 99-109 .

Mikkelson, Gregory M., "On the Intrinsic Value of Everything," *Environmental Ethics*, à paraître.

Mikkelson, Gregory M., "Weighing Species," *Environmental Ethics*, vol. 33, no. 2, 2011a, pp. 185-196.

Mikkelson, Gregory M., "Diversity and the Good," in de Laplante, Kevin, Brown, Bryson et Kent A. Peacock (eds.), *Philosophy of Ecology*, Amsterdam, North-Holland, 2011b, p. 399.

Mikkelson, Gregory M. et C. A. Champan, "Individualistic Environmental Ethics: A *Reductio Ad Exstinctum*?," *Environmental Ethics*, à paraître.

Miller, Peter, "Value as Richness: Toward a Value Theory for the Expanded Naturalism in Environmental Ethics," *Environmental Ethics*, vol. 4, no. 2, 1982, pp. 101-114.

Norton, Bryan G., "Why I Am Not a Nonanthropocentrist: Callicott and the Failure of Monistic Inherentism," *Environmental Ethics*, vol. 17, no. 4, 1995, pp. 341-358.

Nozick, Robert, *Philosophical explanations*, Cambridge, Harvard University Press, 1981.

O'Neill, John, "Varieties of Intrinsic Value," *Monist*, vol. 75, no. 2, 1992, pp. 119-137.

Osborne, Harold, *Theory of Beauty: an Introduction to Aesthetics*, New York, Philosophical Library, 1953.

Rabinowicz, Wlodek, "Value Relations," *Theoria*, vol. 74, no. 1, 2008, pp. 18-49.

Regan, Tom, *The Case for Animal Rights*, Berkeley, University of California Press, 1983.

Rolston, Holmes, "Values In Nature," *Environmental Ethics*, no. 3, 1981, pp. 113-128.

Rolston, Holmes, "Values in Nature and the Value of Nature," in Robin Attfield et Andrew Belsey (eds.), *Philosophy and the Natural Environment*, Cambridge /New York, Cambridge University Press, 1994, pp. 13-31.

Salthe, Stanley N. et Barbara M. Salthe, "Ecosystem Moral Considerability: A Reply to Cahen," *Environmental Ethic*, vol. 11, no. 4, 1989, pp. 355-361.

Stone, Christopher D., *Earth and Other Ethics: the Case for Moral Pluralism*, New York, Harper & Row, 1987.

Stone, Christopher D., "Moral Pluralism and the Course of Environmental Ethics," *Environmental Ethics*, vol. 10, no. 2, 1988, pp. 139-154.

Taylor, Paul W., "The Ethics of Respect for Nature," *Environmental Ethics*, vol. 3, no. 3, 1981, pp. 197-218.

Taylor, Paul W., *Respect for Nature: A Theory of Environmental Ethics*, Princeton, Princeton University Press, 1986.

Thompson, Janna, "A Refutation of Environmental Ethics," *Environmental Ethics*, vol. 12, no. 2, 1990, pp. 147-160.

Varner, Gary E., "Biological Functions and Biological Interests," *Southern Journal of Philosophy*, vol. 28, no. 2, 1990, pp. 251-270.

Varner, Gary E., "No Holism Without Pluralism," *Environmental Ethics*, vol. 13, no. 2, 1991, pp. 173-179.

Varner, Gary E., *In Nature's Interests?: Interests, Animal Rights, and Environmental Ethics*, Oxford, Oxford University Press, 1998.

THINKING GLOBALLY, ACTING LOCALLY: PARTIALITY, PREFERENCES AND PERSPECTIVE

GRAHAM ODDIE
UNIVERSITY OF COLORADO

ABSTRACT:

A rather promising value theory for environmental philosophers combines the well-known fitting attitude (FA) account of value with the rather less well-known account of value as *richness*. If the value of an entity is proportional to its degree of richness (which has been cashed out in terms of *unified complexity* and *organic unity*), then since natural entities, such as species or ecosystems, exhibit varying degrees of richness quite independently of what we happen to feel about them, they also possess differing degrees of mind-independent and subject-independent value. In particular, their value is not dependent on the desires or preferences of humans. The fitting attitudes account of value, at least as it is standardly developed, demands isomorphic evaluative responses on the part of all valuers. In particular, it entails that all valuers should have isomorphic preferences. But this seems absurd. I consider three different strategies with which the fitting attitude theorist can deflect this challenge. The first makes use of an account of non-standard value relations in terms of permissible preference orderings. The second appeals to value appearances and the associated notions of value distance and value perspective. The third involves an account of the ultimate bearers of value as properties, rather than as propositions or states of affairs. These strategies are not all mutually incompatible. While it isn't possible to combine the first and second strategies, it is possible to combine the first and third strategies, and also to combine the second and the third.

RÉSUMÉ :

Pour les philosophes de l'environnement, une théorie des valeurs plutôt prometteuse allie l'analyse de la valeur en termes d'attitudes appropriées, déjà bien connue, à une analyse moins bien connue, celle de la valeur en termes de *richesse*. Si la valeur d'une entité est proportionnelle à son degré de richesse (celle-ci étant rendue en termes de *complexité unifiée* et d'*unité organique*), étant donné que les entités naturelles, comme les espèces ou les écosystèmes, présentent des degrés de richesse variables aucunement liés à ce que nous pourrions ressentir à leur égard, elles possèdent également divers degrés de valeur indépendante du sujet et de la pensée. En particulier, leur valeur ne dépend ni des désirs ni des préférences des êtres humains. L'analyse de la valeur en termes d'attitudes appropriées, tout au moins selon son élaboration habituelle, exige des réponses évaluatives isomorphes de la part de toute entité attribuant une valeur. Or, cela suppose que ces dernières doivent posséder des préférences isomorphes, ce qui nous semble absurde. Nous examinons trois différentes stratégies que peut adopter le théoricien de l'attitude appropriée pour faire face à ce défi. La première met à profit une analyse non standard des relations de valeur en termes de classements de préférence permis. La deuxième fait appel aux apparences de valeur et aux concepts connexes de distance de valeur et de perspective de valeur. La troisième fait recours à l'analyse des porteurs ultimes de valeur en tant que propriétés, et non comme propositions ou états de fait. Ces trois stratégies ne sont pas toutes mutuellement incompatibles. Bien qu'il soit impossible de combiner la première et la deuxième stratégie, il est possible d'unir la troisième à la première et à la deuxième.

FITTING ATTITUDES AND VALUE AS RICHNESS

A fitting attitude (FA) account of value deems something valuable just to the extent that it is appropriate, or fitting, or obligatory to take a certain attitude towards it. The account of value as richness holds that the value of an entity is proportional to its degree of *richness* (or its *unified complexity* or its *organic unity*). The combination of these two theses yields a promising value theory for environment philosophers. Natural entities, such as species or ecosystems, exhibit varying degrees of richness, and they do so quite independently of what we happen to feel about them. So, if value is richness, these entities also possess differing amounts of mind-independent and subject-independent value. Their value is not dependent on the desires or preferences of humans. And given the fitting attitudes account, their value makes certain affective and conative responses on the part of humans and other receptive beings fitting, appropriate, or obligatory. So this combination entails that whatever our actual responses to the natural world happen to be, it is fitting to embrace, preserve, pursue and enjoy the value of those rich systems that populate it¹.

However, the fitting attitude account of value faces a partiality problem. The account, at least as it is standardly developed, demands isomorphic evaluative responses from all valuers. If something is good, say, then every valuer should favour it. And if one thing is better than another, then every valuer should prefer the former to the latter. The value-responses of all valuers should thus be isomorphic to the value facts and hence to each other. Consequently, if two valuers have differing preferences, then at least one of them must be responding unfittingly. But that, it seems to me, is quite incredible. There appear, to me at least, to be large areas over which responses to value not only do, but should vary from one valuer to another.

I will consider three different ways for the fitting attitude theorist to avoid the isomorphic response challenge. The first makes use of Rabinowicz's account of non-standard value relations in terms of permissible preference orderings. The second appeals to value appearances and the associated notions of value distance and value perspective. The third involves an account of the ultimate bearers of value as properties, rather than as propositions or states of affairs. These different strategies are not all incompatible, and in the end, the fitting attitude theorist may have to invoke more than one of them in order to accommodate apparently legitimate variations in our preferences and responses to value.

FITTING ATTITUDES AND THE ISOMORPHIC RESPONSE OBJECTION

The fitting attitude schema ties values to attitudes in a systematically pleasing way. At first blush, the schema seems rather obvious in the case of certain thick evaluative attributes: the admirable is what it is fitting or appropriate to admire, the revolting what it is fitting or appropriate to be revolted by, and so on. But it is also at least plausible for the thin evaluative attributes: goodness seems closely connected to what is desirable, and the desirable is what it is fitting to desire².

If the FA schema is applicable to all values, then it holds out the promise of a reduction of values to their associated attitudes together with the notion of fittingness. Where *favouring* and *disfavouring* are placeholders for the fitting response to bearers of V , the fitting attitude schema is this:

X has value/disvalue V if and only if it is fitting to favour/disfavour X .

A fitting attitude reductionist would strengthen the schema: entity X has value/disvalue V if and only if *and because* it is fitting to favour/disfavour X . This strengthening is intended to capture the idea that the value is possessed by the entity *in virtue of* the fittingness of the attitude, not vice versa. But this reductionist bid founders on the well-known and much discussed Wrong Kinds of Reasons objection³.

Suppose we begin with a deontic notion of fittingness—i.e., favouring is fitting if favouring is the attitude that you *ought* to have or adopt. Now suppose an evil demon threatens the world with something terrible unless you admire him for making that very threat. Then you ought to admire him, even though he isn't in fact admirable⁴. I will come back to what I take to be the fitting attitude theorist's best response to the WRK problem; what I want to examine first is a much less discussed objection to the FA schema.

The basic FA schema matches up values with attitudes. As standardly articulated, the fittingness of a response is not supposed to be a subject-relative matter. It is fitting for one to favour X if and only if it is fitting for everyone to do so. It follows that the responses of any two valuers who are responding fittingly (call these "fitting responders") are isomorphic with the value facts and hence isomorphic with each other. Now suppose, as seems plausible, that desire is the fitting response to the desirable (the good) and preference is the fitting response to betterness. If one valuer prefers X to Y and another prefers Y to X , then one or the other (or possibly both) must be responding unfittingly. But that seems a tad counterintuitive. Surely there are some occasions that invite a range of different preferences, all of which are fitting responses to value. Different valuers can, quite legitimately, respond in different ways to things which possess their values independently of those valuers' responses.

Here is an example. Consider Erin. Erin loves egrets. She has always loved egrets. Egrets are her passion. She grew up near some wetlands (hereafter "the Wetlands"), spending all the spare time in her childhood mucking about in them—studying, photographing, drawing and monitoring the egrets and the other wildlife there—and she still does this whenever she can. She studied evolutionary biology at college, but after that went into a career in equity, and now has a lucrative day job working for a hedge fund. The Wetlands have recently been threatened by a housing and supermarket development project that will completely destroy the egrets' habitat along with the rest of the rich ecosystem in which they thrive. Distressingly for Erin, it is the very hedge fund for which she works that has a controlling share in the project. Erin is deeply conflicted—she

loves her job otherwise and makes a huge amount of money—but eventually she decides to leave the hedge fund behind, to take much less well remunerated part-time work, and to spend her own time founding the Wetlands Preservation Project to fight the development.

Tammy, some ten thousand miles away, loves the Takahe, a native bird of New Zealand. Only four were observed in the nineteenth century, and by 1930 they were believed extinct. But although they came perilously close, they were not in fact extinct. After a sighting in 1948 in Fiordland, some modest preservation strategies were implemented, and now there are roughly two hundred birds. Tammy is keen to ensure the continuation of the Takahe in their habitat, but she has no particular aptitude for field work or environmental activism and decides that the best way for her to promote the survival of the Takahe is to work harder at her lucrative career in real estate and devote all the extra cash she earns to the Takahe Preservation Foundation.

Now suppose Erin also learns about the Takahe project. And suppose she has an opportunity to re-join the hedge fund, where her apostasy would be forgiven. She could go back, abandon the Wetlands to her firm's development project, and send all the extra money she makes to the Takahe Preservation Foundation. But she just has no real desire to do that, so she sticks with her modest part-time work and continues pouring her energies into the Wetlands Preservation Project.

Tammy and Erin appear to have different desires and preferences. Erin prefers doing hands-on work with the Wetlands project to working harder in her lucrative career and donating cash to the Takahe project. Tammy, by contrast, prefers to work harder at her lucrative career in real estate and donate her spare cash to the Takahe Foundation. Do these differences entail that at least one of these two preference structures is unfitting? Is Erin's pursuit of the Wetlands preservation project together with her decision not to return to her lucrative job and pursue the Takahe preservation project less fitting than Tammy's opposite and equally enthusiastic response? Is Erin's continuing commitment to the Wetlands project—even after she learns of the Takahe project—misguided? Or is Tammy's response defective? Should she get out of real estate and, like Erin, pitch in at the ground level?

Erin and Tammy's different ways of responding to these values both seem quite legitimate or fitting. Any value theory stating that at least one of these two valuers must be responding unfittingly—on the grounds that, to respond fittingly, both must exhibit isomorphic responses to the same value bearers—seems wrong. In particular, a value theory that implies that Erin can only respond fittingly by changing her preferences in some way—by abandoning her desire to work on the Wetlands project, to work harder at her job on Wall Street, and devote her excess monetary resources to the Takahe project—seems highly questionable, if not absurd.

VALUE AS RICHNESS

On the face of it, there are very different kinds of valuable things: happiness, desire satisfaction, pleasure, knowledge, virtues, works of art, music, persons, justice, living organisms, species and ecosystems, to name a few. Recently, Chris Kelly (2003 and 2014) has argued that a certain theory of value can capture most of these very diverse value judgments⁵. It is the thesis that value just is, or is proportional to, the *richness*, or *unified complexity*, or *organic unity* of the value bearer.

Value as Richness (VR)

Degree of value is proportional to degree of richness (or unified complexity, or organic unity).

This is a monistic theory of value. It has had a small number of subscribers in the history of philosophy (most notably Leibniz and Nozick⁶), but it is neither well known nor fully developed yet. It promises to capture, unify and explain a number of otherwise disparate intuitions about value. In particular, it makes rather good sense of the judgments that many environmental philosophers have urged concerning the mind- and subject-independent value of entities such as ecosystems. Indeed, ecosystems are paradigm exemplars of rich entities, for they exhibit high degrees of unified complexity or organic unity.

As a monistic theory value, VR is compatible with a strong form of value comparability.

Strong Comparability

Any two entities (of the same ontological type) are strongly comparable with respect to their degree of unified complexity (or degree of richness): either *X* is richer than *Y*, or *Y* is richer than *X*, or *X* and *Y* are equally rich.

Now suppose, as also seems plausible, that the fitting response to betterness is *preference*. Then we have:

Preferences should track richness.

One ought to prefer *X* to *Y* if and only if *X* is richer than *Y*.

Consider then the relative richness of Erin's Wetlands ecosystem and Tammy's Takahe ecosystem. Assuming strong comparability, there is some fact of the matter about which of the two is richer. If the Takahe ecosystem is richer, then both Erin and Tammy ought to prefer the preservation of the Takahe ecosystem to the Wetlands ecosystem (all other things being equal). If the two ecosystems have an exactly equal degree of richness, then Erin and Tammy should both be indifferent between them. Presumably they should also be indifferent as to which one to commit resources to, deciding perhaps by the toss of a coin. If the two ecosystems are of exactly equal richness, then any tiny increase in the richness

of one ecosystem should move both conservationists to line up behind that preservation project. So, for example, a tiny increase in the value of pursuing the Takahe ecosystem should be enough to move Erin to abandon her commitment to the Wetlands ecosystem. But this seems odd. Any account of the fittingness of responses should be able to accommodate diverse legitimate responses, including those exhibited by Tammy and Erin in the scenarios above.

Suppose that the preservation of the Takahe ecosystem would make the world richer than would the preservation of Erin's Wetlands. This seems plausible: there are plenty of good egret habitat available worldwide. Even if the development of Erin's Wetlands would eliminate the egret ecosystem from Erin's particular locality, egrets and their supporting ecosystems are under no particular threat globally. The failure of Tammy's Takahe preservation project, on the other hand, would seriously diminish the richness of the world. So, if preferences ought to track richness, then Erin ought to prefer the success of the Takahe preservation project to the success of her Wetlands preservation project. Presumably, a consequence of this theory is that she ought to want to increase her hours in the odious Wall Street hedge fund (the one that is destroying her beloved Wetlands), and send all her spare cash to the Takahe preservation project. But she doesn't have those desires and preferences. She stubbornly prefers to promote the poorer outcome (the preservation of the Wetlands ecosystem) to the richer outcome (the preservation of the Takahe ecosystem). So she is an unfitting responder. One could, of course, simply bite the bullet here and embrace the conclusion that all responders should have isomorphic preferences. But that seems rather extreme, so here I consider three possible ways for the proponent of FA to resist⁷.

PARITY AND PERMISSIBLE PREFERENCES

In his "Value relations" (and in a number of subsequent papers), Wlodek Rabinowicz develops a normative (or deontic) notion of fittingness within a general fitting attitude account of value relations:

X is more valuable than *Y* if and only if it is *deontically* fitting to prefer *X* to *Y*⁸.

He notes that there are two different normative, or deontic, notions of fittingness—a strong notion and a weak notion. A preference might be *obligatory* (the strong notion of fittingness) or it might be *permissible* (the weak notion of fittingness).

Weak deontic fittingness

It is weakly fitting to prefer *X* to *Y* if and only if it is *permissible* to prefer *X* to *Y*.

Strong deontic fittingness

It is strongly fitting to prefer *X* to *Y* if and only if one *ought* to prefer *X* to *Y*.

Rabinowicz formally models the permissibility and obligatoriness of particular preferences in terms of a collection of rationally permissible preference orderings.

It is a common assumption in value theory that both rational preferences and betterness induce complete orderings in their domains. This seems rather a tall order. It would be more reasonable to countenance some gaps in the orderings—that is, to countenance *partial* orderings of the elements in the domain to be compared for value. Allowing partial orderings won't by itself eliminate the isomorphic response objection. For if there is only one permissible partial ordering then only one preference ordering is permissible (albeit an ordering with gaps). And, by the FA schema, that preference ordering would be isomorphic to the betterness relation. In order to avoid the isomorphic response objection, the FA-theorist must therefore countenance more than one permissible value ordering (with or without value gaps in it). Allowing more than one permissible preference ordering yields a certain latitude in what preferences are permissible—the larger the class, the greater the latitude.

Indeed, the model was developed at least in part to accommodate Chang's non-standard value relation of *parity* within a fitting attitude approach. *X* and *Y* are on a par if they are in some sense comparable for value, but none of the standard value relations hold (namely: *X* is better than *Y*; *Y* is better than *X*; *X* has exactly the same value as *Y*). Chang appeals to the *small improvement argument* to motivate the possibility of parity. For example, consider two ecosystems that seem to be in the same ballpark of value: Erin's Wetlands and, say, a comparably rich grassland ecosystem in Colorado. Is one more valuable than the other? One might be reluctant to subscribe to either judgment. There are of course epistemic problems involved here. There are many elements of both ecosystems to evaluate and weigh, and one might overlook some of them or make a wrong judgment. There is also the problem of how to combine the value of the individual components into an overall evaluation of the ecosystem. If the ecosystem is valuable in proportion to its richness (or organic unity), then its value clearly will not be the sum of the values of individual components. But the problem might be deeper than these epistemic factors. It could be that neither ecosystem is in fact superior to the other, yet it does not follow that they must then be either of equal value or else incomparable for value.

Suppose one grants that neither ecosystem is in fact more valuable than the other. Since richness involves at least two dimensions (unity and complexity), given VR it is possible that neither is clearly better than the other⁹. The Grasslands might exhibit greater unity, while the Wetlands exhibit greater complexity, say, and it may not be determinate how to weight these two factors in an overall assessment of richness. Now consider a small improvement to Erin's Wetlands. Suppose a small quantity of some mildly harmful chemical has been slowly leaching into the Wetlands from a nearby industrial plant, and Erin manages to get this source of contamination sealed off, thereby enhancing the unity of the Wetlands. Call this improved wetland ecosystem Wetlands⁺. Wetlands⁺ is clearly more valuable than Wetlands. And if Wetlands and Grasslands are equal in value, then Wetlands⁺ would also be clearly better than Grasslands. But while Wetlands⁺ is clear-

ly better than Wetlands, that does not seem to make Wetlands⁺ clearly better than Grasslands. So Wetlands and Grasslands aren't equal in value either. But neither are Wetlands and Grasslands totally incomparable for value. They are clearly in the same ballpark. For example, if an environmental agency had a limited amount of money to spend on a substantial improvement to one ecosystem, both a Wetlands proposal and a Grasslands proposal might legitimately make it into the final round, ahead of various other clearly inferior proposals. If Wetlands won and was *substantially* improved thereby (call that Wetlands⁺⁺), then the result might well be richer, and hence more valuable, than Grasslands.

Betterness is given by strong fittingness of preference (one *ought* to prefer X to Y) which holds just in case X is preferred to Y in every permissible preference ordering. X and Y are *on a par* provided that it is (weakly) fitting to prefer X to Y but that it is also (weakly) fitting to prefer Y to X . That is, in some permissible preference orderings, X is preferred to Y , while in others, Y is preferred to X . A small improvement to X (X^+) will be preferred to X in every preference ranking (i.e., X^+ is strictly better than X), but if the improvement is small enough, then X^+ and Y , like X and Y , will also be on a par. However, a substantial enough improvement in X (X^{++}) would be better than Y —i.e., X^{++} is preferred to Y in every permissible preference ranking.

Suppose, then, that we list a number of permissible (kinds of) preference ranking involving our various ecosystems:

Preference ranking A: Takahe $>_A$ Wetlands⁺⁺ $>_A$ Wetlands⁺ $>_A$ Wetlands $>_A$ Grasslands

Preference ranking B: Takahe $>_B$ Wetlands⁺⁺ $>_B$ Wetlands⁺ $>_B$ Grasslands $>_B$ Wetlands

Preference ranking C: Takahe $>_C$ Wetlands⁺⁺ $>_C$ Wetlands⁺ $>_C$ Wetlands \approx_C Grasslands

Preference ranking D: Takahe $>_D$ Wetlands⁺⁺ $>_D$ Wetlands⁺ \approx_D Grasslands $>_D$ Wetlands

.....

In every permissible preference ranking: Takahe is preferred to Wetlands and Takahe is preferred to Grasslands. Also Wetlands⁺⁺ is preferred to Wetlands⁺, which in turn is preferred to Wetlands. Where $>_V$ is the relation of *greater value*, the intersection of all the permissible preference rankings, we have:

Takahe $>_V$ Wetlands⁺⁺
 Wetlands⁺⁺ $>_V$ Wetlands⁺ $>_V$ Wetlands
 Takahe $>_V$ Grasslands.

But none of the following relations hold:

Wetlands $>_V$ Grasslands
 Grasslands $>_V$ Wetlands;
 Wetlands \approx_V Grasslands.

Wetlands and Grasslands are on a par given this collection of admissible preference rankings. Rabinowicz's model thus allows for the fittingness of a preference for Wetlands over Grasslands, or for Grasslands over Wetlands, as well as indifference between the two. All these are fitting preferences—that is, they are all *weakly* fitting. So not all fitting preferences have to be isomorphic either to actual value or to each other.

Despite the latitude that Rabinowicz's model allows, it cannot spare Erin condemnation for her preferring the Wetlands project to the Takahe project. The model cannot countenance the compatibility of Takahe $>_V$ Wetlands with the permissibility of a preference for Wetlands preservation over Takahe preservation. But it is worse than this. Suppose that the Grasslands preservation is neither better nor worse than Wetlands preservation, but exactly the same value. This means that in every permissible preference ranking they are equi-preferred. So there is no permissible preference ranking in which Wetlands preservation is preferred to Grasslands preservation. If this is the case, then Erin's pursuing the Wetlands' preservation over the Grasslands' preservation is *strictly impermissible*. Although she may toss a coin to decide whether to support one cause or the other, if she does launch herself whole-heartedly and passionately into the Wetlands preservation without so much as a second thought for the equally valuable Grasslands and the value of their preservation, then her desires are to that extent defective.

REPRESENTATION, PERSPECTIVE AND DISTANCE

If fittingness is a deontic notion, a matter of what one *ought* to prefer, then the *Wrong Kinds of Reason* objection looms¹⁰. Suppose a powerful but evil demon threatens to bring about the very worst outcome unless you admire him, and will spare the world the worst outcome only if you do so admire him. Then you ought to admire him—even though he is not in the least admirable. Similarly, if the demon threatens to bring about the very worst outcome unless you prefer that outcome to all others, and will spare the worst outcome if you do so prefer it, then you ought to prefer it—but the outcome you ought to prefer is not good. The same kind of counterexample can be made to work against an axiological notion of fittingness as well (i.e., that the good just is what it is good to favour).

Various attempts, mostly rather ad hoc and unsatisfactory, have been made to square these normative readings of fittingness with WKR counterexamples. I will not review those here. But an FA theorist can embrace an alternative notion of fittingness that sidesteps the WKR counterexamples altogether. Tappolet calls it a *descriptive* reading, as opposed to a *normative* reading. She motivates this notion by appealing to a perceptual theory of emotions, whereby an emotion involves a non-conceptual representation of value:

An alternative conception [...] is that the appropriateness of emotions is a matter of representing things as they are. In the relevant sense, appropriate emotions are emotions that are correct from an epistemic point of view¹¹.

I will call this the *representational* (rather than descriptive) notion of fittingness. Where A^V is the attitude that pairs with value V , the FA schema becomes:

X is V if and only if it is representationally fitting for one to take attitude A^V to X .

Suppose that attitude A^V consists in the object X 's *appearing* to have value V . For an attitude A^V to X to be representationally fitting is for A^V to represent X as possessing the value V , and for X to in fact possess the associated value V . For example, for you to admire the aforementioned demon, the demon must appear in some way admirable to you. And it is fitting to admire the demon just in case that appearance represents the demon accurately. Representational FA will not, of course, underwrite a *reduction* of value to non-evaluative concepts, since on this account, the fitting attitudes themselves must have evaluative representational content (i.e., they are representations of things as possessing value). But if one is not intent on reducing goodness to some other properties, then this is not an objection to the representational reading of fittingness as such. A fitting attitude theorist can endorse the schema without touting it as a reductive tool.

The Representational FA theorist can easily avoid the demon objection and others like it. She can say: "It isn't fitting to admire the demon, for that attitude would be a grossly inaccurate representation. It would be to represent the demon as admirable, when he is in fact contemptible. But the best thing to do, all things considered, is to avert the demon's threat of harm, even though that involves taking an attitude towards the demon that clearly doesn't represent him as he is. In order to do the right (or best) thing, you simply have to suppress your (perfectly fitting) contempt for the demon and do your best to admire him." Similarly, even if desire is the right or best attitude to take to the worst outcome in light of the demon's threat, that does not make desire a *fitting* response to the worst outcome *per se*.

How does the representational FA schema bear on the problem of isomorphic preferences? Augustine articulated an early version of what is sometimes called the *guise of the good*:

[...] [In] the [pull] of the will and of love, appears the worth of everything to be sought or to be avoided, to be esteemed of greater or lesser value¹².

I call this the *value appearance* thesis. It has a number of related components. First, there are appearances of value: these are perception-like rather than belief-like (something can *appear* a certain way without one's believing it to *be* that way.) Second, a desire for X is or involves an appearance of the goodness of X (and a preference for X over Y is or involves an appearance of X 's possessing more value than Y .) Third, an appearance of value is fitting just in case it is an accurate representation of its object.

The isomorphic response thesis would accordingly take the following form:

X is better than *Y* if and only if it is fitting for *X* to appear better than *Y* (if and only if it is fitting to prefer *X* to *Y*). So the preferences of any two fitting responders will be isomorphic.

The analogue of this idea for regular perception would be patently absurd. It is not a requirement of accurate perception that for any two observers the appearances should be isomorphic to reality and hence isomorphic to each other. This is because perceptual representations are representations *from a certain point of view*, and as such they legitimately incorporate perspectival effects. Suppose two objects *X* and *Y* are exactly the same size. Dom is closer to *X* than to *Y* while Eric is closer to *Y* than to *X*. To Dom, *X* will appear larger than *Y*, while to Eric, *Y* will appear larger than *X*. But neither response is an unfitting or inaccurate perceptual representation of the objects. Rather, it is fitting that the objects should appear differently to each perceiver given the different relations that the former bear to the latter. Perception is always perception of objects as they stand in relation to the perceiver. From the Earth, the setting moon should appear larger than the rising sun, even though it is in reality much smaller. As such, the appearances for two differently situated perceivers should not in general be isomorphic to reality or to each other.

If there are genuine appearances of value, then there should also be value analogues of distance and perspective. If we suppose that desires and preferences value appearances, then they should depend not just on the evaluative properties of the objects of those appearances but also on how one stands in relation to those objects. There should be a perspectival element to desire, just as there is a perspectival element to visual perception. Just as it is entirely fitting for distant objects to appear small by comparison with nearby objects, states that are far from you in value space should not loom as large in your desires and preferences as states that are close to you. The weals and woes of those who are closely connected to you, those for whom you care deeply, are closer to you in value space than the weals and woes of distant strangers. Suppose, for example, that pain is bad, and that qualitatively identical pains have the same disvalue whoever is the subject of that pain. Suppose that your daughter and some stranger are experiencing extreme and qualitatively identical pains. You are far more averse to your daughter's being in extreme pain than to the stranger's being in pain. Of course you are also averse to the stranger's predicament, but if there were only one dose of morphine at hand, then you would not find yourself completely indifferent between the stranger's getting that dose and your daughter's getting it. But if we take a deontic approach to fittingness, then it is clear that you *ought* to be completely indifferent as to whether the last dose of morphine goes to your daughter or to the stranger. That seems odd. I think that it is completely appropriate or fitting for you to care more about the pain of people you love and with whom you have deep connections than about the pain of people with whom you have no connections at all—which is not to say that it would be fitting not to

care at all about the stranger's pain, nor that you should believe that your daughter's pain is worse. I mean simply that it is appropriate for you to prefer that your daughter receive the last dose of morphine. If desires and preferences are non-doxastic appearances of value then this falls out rather naturally.

We can now apply this perspectival analysis to the conception of value as richness. Some rich states of affairs are much closer in value space to a particular value responder than other equally rich states. The survival of the Wetlands is much closer in value space to Erin than the survival of the Grasslands, even though the two ecosystems are (let's suppose) equally rich and thus equally valuable states of affairs¹³. Even though they are equally valuable, it is entirely fitting that the Wetlands should loom larger in Erin's preferences and desires (that is, in how valuable they appear to her) than the Grasslands. If it is entirely fitting for Erin to prefer the preservation of the Wetlands to the preservation of the Grasslands, then it is also fitting for her to be more motivated to promote the preservation of the Wetlands over that of the Grasslands. And if she acts on such fitting motivations, then presumably that is something for which we cannot fittingly criticize or condemn her.

One might object that this is all well and good in states of affairs that are of *exactly* equal value, or even perhaps for states that are on a par. In Erin's case, her closer connection to the Wetlands can legitimately tip the balance only because there is either a value tie or because there is parity. But if one state of affairs is more valuable than another, then such connections become completely irrelevant. The only fitting response is to prefer the more valuable state to the less valuable one. But if distance can make a difference to legitimate preferences in the case of exact equality of value, then it would constitute an inexplicable discontinuity if it could not make any difference at all in a case of two states that are extremely close in value but not quite equal.

PROPERTIES AS VALUE BEARERS

There are two problems with the perspectival reply to the isomorphic response objection. First, the notion of value perspective and distance in value space, as it stands, is somewhat metaphorical. Is there any way of making it less metaphorical, of theoretically cashing out the notion of perspective? Second, it is not clear why desire and preference should be fitting responses to the (distance-discounted) richness of states of affairs. Why should the richness of states of affairs present itself to a value responder as something that it is fitting to desire?

The value appearance thesis is that desires and preferences are appearances of value. (It is *not* the thesis that all appearances of value are desires.) For the value appearance thesis to be on the cards, the objects of desire must also be bearers of value. The standard view is that desire and preference, like belief, are *propositional* attitudes, and that the objects of such attitudes are either propositions or closely related entities that propositions pick out, namely states of affairs. This dovetails nicely with a widespread monistic view about value bearers, namely that propositions or states of affairs are the fundamental bearers of intrinsic

value¹⁴. If other things—such as individuals or properties—have value, then they have it in virtue of the intrinsic value of the states of affairs in which they feature.

The propositional attitude view sits happily with the surface grammar of some desire claims, but not of all. Consider:

Oliver desires the Olympic gold medal for the 15km cross-country skiing event.

Harvey wants a hokey-pokey ice-cream.

Erin prefers the Wetlands to the Grasslands.

An Olympic gold medal, a hokey pokey ice cream, Wetlands and Grasslands all appear to function here as objects of desire, or of preference, but they are not obviously propositions or states of affairs. However, whenever a desire is specified without specifying an explicit propositional object, it is usually easy to find an essentially equivalent and natural specification of the desire that makes a propositional analysis plausible. It is not too farfetched to say that Oliver's desire for *the Olympic gold medal*, or Harvey's desire for *a hokey-pokey ice-cream* are desires for certain interactions with the entity at issue: to *win the gold* or to *eat a hokey-pokey ice cream*. What Erin prefers is *preserving the Wetlands* to *preserving the Grasslands*, or *working on the Wetlands preservation project* to *working on the Grasslands preservation project*. Finally, Oliver's desire *to win the Olympic gold* will be further parsed as Oliver's wanting it to be the case *that Oliver wins an Olympic gold*. But what state does Erin prefer to what? Erin prefers *that Erin preserve the Wetlands* to *that Erin preserves the Grasslands*. This last formulation may sound a bit strange, but one can get used to it fairly quickly.

There is, however, a rival to the propositional view that secures the uniformity of the ontological category without any cost to naturalness or flexibility. Suppose Oliver and Orlando are rivals for the Olympic gold in the 15km cross-country skiing event. Oliver says *I really want to win Olympic gold!* Orlando chimes in *So do I!* The common object of their two desires is *to win the Olympic gold medal in the 15km cross-country event* (or *to win the Olympic gold* for short). On the propositional view, they can only want the same thing if they want the very same state of affairs to obtain: but there is no state of affairs that corresponds to their common desire in this case. Rather, what each wants is a *property*, that of *winning the Olympic gold*. In general, what one desires is to possess or to acquire some property, and what one prefers is the having of some property to the having of some other property¹⁵. The property view can make literal sense of the claim that Oliver and Orlando want *the very same thing* (to have the property of *winning the Olympic gold*). If Oliver's desire is satisfied, then Orlando's is frustrated, and vice versa.

In addition to wanting to preserve the Wetlands, Erin also wants *the Wetlands to be preserved*. But *the Wetland's being preserved* is a state of the world, rather than

a property of Erin. So, at the very least, the property view has to be able to capture the class of desires having states of the world as their apparent objects. Happily, there are quite natural property-surrogates for states of the world that can serve as the objects of these apparently state-directed desires. I will call these *global* properties. For each state of the world S there is exactly one global property, $P(S)$, characterized as follows: X has property $P(S)$ if and only if S obtains. $P(S)$ is the property of *its being such that S obtains*¹⁶. Since X doesn't occur on the right-hand side of the definition, $P(S)$ is *global* in the sense that it is had by some individual if and only if it is had by all. Either every individual has $P(S)$ or none have it. The circumstances in which it is had by one and all are just those in which the world is in state S . So wanting to have $P(S)$ is tantamount to wanting S itself to obtain. We can retrieve the class of propositional desires, or state desires, from desires the objects of which are these global properties. When Erin and Tammy both have the apparently state-directed desire *that the Wetlands be preserved*, the common object of their desire is a global property.

Global properties can be contrasted with *local* properties. A local property is a property that can be had by some individuals without being had by all. The property of *preserving the Wetlands* is local, since Erin can have it while Tammy lacks it.

What does desire fulfillment consist in, on the property view? On the state or propositional view, the desire that P is fulfilled if and only if the object of the desire, the proposition P itself, is true. But properties do not bear truth values. Further, on the property view, Oliver and Orlando can share the *very same* desire (*winning the Olympic gold*). But how can those be the *same* desire if the fulfillment of Oliver's desire will constitute the frustration of Orlando's? That is a violation of Leibniz's principle. However, this puzzle is easily dissolved. *Desire* suffers from the usual state/object ambiguity: *Oliver's desire* can pick out the state that consists in Oliver's desiring something, or it can pick out the object of Oliver's desiring. Fulfillment is not solely a feature of the object of desire. For Oliver's desire (to win the Olympic gold) to be fulfilled, *Oliver* (rather than Orlando or anyone else) must acquire the object of his desire. In this case, Oliver must acquire the property of winning the gold.

Consider a selection of local and global properties: winning the Olympic gold; eating a hokey pokey ice-cream; the *Goldberg Variations* being played; hearing the *Goldberg Variations* being played; playing the *Goldberg Variations*; preserving the Wetlands; the Wetlands' being preserved by someone or other. These are all possible objects of desire on the property view, but they are not only objects of desire—they are also apt subjects of value attributions. Plausibly, all these properties have some value, although some are more valuable than others. Eating a hokey pokey ice cream has some minor value; hearing the *Goldberg Variations* being played is more valuable; and playing the *Goldberg Variations* is more valuable still. Winning the Olympic gold may be somewhere in the ballpark of playing the *Goldberg Variations*. So on the property view, it is plausible that the possible objects of desire are also apt bearers of value.

Of course, not all bearers of value are possible objects of desire. The *Goldberg Variations* is a valuable musical work. Even if it turns out to be a property (say,

the property had by all and only possible performances of the work), that is not the kind of property that it makes sense to *desire*. No rational person wants to *instantiate* that property. To so desire would be a rather odd category error, as one can rationally desire only those types of property that one could conceivably instantiate. However, this is no obstacle to the value appearance thesis, since the latter does not require that all bearers of value be possible objects of desire.

What makes this shift in the primary bearers of value—from propositions to properties—a rather promising step in tackling the isomorphic preference problem is this: local properties (unlike propositions or global properties) have their own built-in *perspective* or *point of view*. A local property of individuals does not characterize a state of the world; rather it characterizes a way of *being in the world*.

For example, consider the following four properties, each specifiable in terms of the more basic properties of playing the *Goldberg Variations* and listening to the *Goldberg Variations*: *playing the Goldberg Variations to an audience of at least one* (P); *being a member of an audience listening to the Goldberg Variations being played* (H); *neither playing nor listening to the Goldberg Variations* (N); *its being such that the Goldberg Variations are played by one person and heard by at least one other* (G). P, H and N are local properties, G is global. P, H and N characterize three very different kinds of niche in any world in which G is instantiated. If you are told that you (and hence everyone) will come to have the property G, then you thereby acquire some interesting information about the world and its richness. But you would be none the wiser as to whereabouts in the structure of the world you yourself would end up—you might be a *Goldberg Variations* player, or a *Goldberg Variations* listener, or you might be neither. Each of the local properties P, H and N does not simply characterize the G-world, but rather characterizes locations within the structure of a G-world.

More generally, if we take any individual *X*, we can ask: *What basic monadic properties does X have, and what basic monadic properties does X lack?* There is a collection of complete specifications of basic monadic properties which partitions the set of individuals into mutually exclusive and jointly exhaustive kinds. Call this a *depth-0 kind*. We can go on to ask: *What basic relations does X bear to various depth-0 kinds of individuals?* This will refine the initial partition into a more fine-grained partition. These are depth-1 kinds of individuals. Each individual falls into one and only one of these depth-1 kinds. We can go on increasing the characterization of an individual's properties and relations to other individuals to any depth of specificity we like, and as we go deeper and deeper, we narrow down more precisely the particular kind of niche that a given individual occupies in the world. At an infinite depth, we arrive at near complete specifications of the kinds of loci that individuals can occupy in the web of being¹⁷. These infinitely deep kinds are rather close to Leibniz's complete individual concepts, though without further assumptions one cannot show that the niches so specified can only be occupied by one individual, or that an infinitely deep kind can only be instantiated in one possible world.

THE RICHNESS OF PROPERTIES

According to VR, states are valuable in proportion to their richness. States are rich in virtue of their *structure*. To establish this, it is sufficient to note that two isomorphic states of affairs must enjoy the same degree of richness or unified complexity. The best explanation for this relationship is that it is the *structure* of a state that determines its degree of richness. In other words, a state's degree of richness corresponds to the richness of its structure. But the structure of a state is the common *property* of all the states that are isomorphic to it. Thus, states possess richness in virtue of the richness of the properties that they instantiate. Consequently, if the thesis of value as richness is applicable to states, it is also applicable to properties.

Some properties are clearly richer (exhibit a higher degree of unified complexity) than others. *Being a human being* involves having a richer structure than does *being an amoeba*. And *being an amoeba* involves having a richer structure than does *being a disorderly heap of stones*. Moreover, these judgments of relative richness accord with the correlative judgments of relative value. Going back to our examples in the previous section, eating a hokey pokey ice cream has a fairly low (albeit non-negative) degree of richness, while hearing and appreciating the *Goldberg Variations* has a greater degree of richness, and playing the *Goldberg Variations* with appropriate skill and appreciation has greater richness still.

Let E be any ecosystem, and consider two local properties and one global property associated with E . (Here D indicates *direct participation*; I indicates *indirect participation* and N indicates *non participation*.)

E^{Dir} : To preserve ecosystem E , to which one is already richly connected through past interactions, by resigning from one's lucrative job and directly devoting one's time and energy to preservation activities.

E^{Ind} : To preserve ecosystem E by working harder at one's lucrative job and donating one's spare cash to E 's preservation.

E^{Non} : To be such that the E 's ecosystem gets to be preserved.

(Note that property E^{Dir} builds in the individual's existing rich connections to the ecosystem E . The other two properties do not, but neither do they exclude that possibility.)

Erin wants to give up her job at the hedge fund and start the Wetlands Preservation Project. Abbreviating Wetlands to \mathbf{W} , Erin *prefers* \mathbf{W}^{Dir} to \mathbf{W}^{Ind} and she *prefers* \mathbf{W}^{Ind} to \mathbf{W}^{Non} . To Erin, it thus seems more valuable to be directly involved in preserving the Wetlands than to be only indirectly involved (i.e., by staying on

at the hedge fund and donating cash to the cause), which in turn seems more valuable than the Wetlands's being preserved by someone or other. Assuming that value is richness, these appearances could be entirely fitting. \mathbf{W}^{Dir} is (plausibly) a richer property to have than either \mathbf{W}^{Ind} or \mathbf{W}^{Non} . If the Wetlands are preserved by someone or other, then *everyone*, no matter what relations they individually bear to the Wetlands, has the global property \mathbf{W}^{Non} . For Erin to have \mathbf{W}^{Non} , the Wetlands must be preserved, but Erin does not have to bear any other connections to them. But to have \mathbf{W}^{Dir} , Erin must not only have the global property \mathbf{W}^{Non} , but in addition she herself must already have various rich connections with the Wetlands and also bear interesting new causal and intentional relations to their preservation. In particular, if she possesses \mathbf{W}^{Dir} , she will be integral to the complexity of Wetlands in a way that she may well not be if someone else preserves the Wetlands courtesy of infusions of her cash.

We can motivate this idea by comparing the relative richness of another pair of properties: playing a rich piece of music, like the *Goldberg Variations*, and paying someone else to play it for you. Suppose you love the *Goldberg Variations*. To hear the *Goldberg Variations* being played with appreciation is a property with a rather high degree of richness. The *Goldberg Variations* themselves have a high degree of unified complexity, of course, and whoever hears and appreciates them gets to participate in that richness. Paying to have someone else play them so that you can hear them is quite a rich property to have. But one who both appreciates *and* skilfully plays the *Goldberg Variations* not only gets to hear them played: other things being equal, she is much more closely tied to that complex unified structure than is the person who merely pays to have them played, by virtue of her being the direct causal conduit through which that structure becomes realized in the world¹⁸. The mental states of the listener and the player both add richness to the original richness of the work itself, but the player's mental state is more closely united with the work than is the listener's state. So the appreciative and skilful player of the *Goldberg Variations* possesses a richer property than the appreciative but passive listener.

If this analysis is right, it suggests that Erin, by participating directly in the preservation of the ecosystem to which she is already richly connected, would become richer than if she simply donated the cash to someone else to do the preserving of the Wetlands (or of some equally valuable ecosystem to which she bears no connection) in her stead. And that in turn is richer than the Wetlands' getting preserved somehow or other. So:

$$\mathbf{W}^{\text{Dir}} >_{\mathbf{V}} \mathbf{W}^{\text{Ind}} >_{\mathbf{V}} \mathbf{W}^{\text{Non}}.$$

Now suppose the Grasslands (\mathbf{G}) have exactly the *same* degree of richness as the Wetlands. Then even though $\mathbf{W}^{\text{Non}} \approx_{\mathbf{V}} \mathbf{G}^{\text{Non}}$, we have that $\mathbf{W}^{\text{Dir}} >_{\mathbf{V}} \mathbf{W}^{\text{Ind}} >_{\mathbf{V}} \mathbf{G}^{\text{Non}}$. So engaging directly or indirectly in the preservation of an ecosystem will be a richer property than its being such that an equally valuable ecosystem is preserved somehow or other. Now consider a slightly richer Grasslands, \mathbf{G}^+ . If

\mathbf{W}^{Ind} is better than \mathbf{G}^{Non} , then \mathbf{W}^{Dir} might well be better than $(\mathbf{G}^+)^{\text{Non}}$, since the added richness involved in one's directly preserving the Wetlands could exceed the added richness involved in Grasslands⁺. This means that there is no obstacle in principle to the possibility that \mathbf{W}^{Dir} is more valuable even than \mathbf{T}^{Non} (where Takahe = \mathbf{T}). It all depends on the differences in richness involved. Of course, the Takahe ecosystem might be so much richer than the Wetlands that the additional richness in the property \mathbf{W}^{Dir} is swamped by the extra richness of \mathbf{T}^{Non} . In that case, Erin's closeness to and connection with the Wetlands cannot make it fitting for her to prefer directly preserving those to indirectly preserving a much richer ecosystem with which she has no pre-existing deep connection.

Finally, even if participating indirectly in the preservation of the Takahe ecosystem (by staying at her hedge fund and donating the money to the Takahe Preservation Fund) is more valuable than its being such that the Takahe ecosystem is preserved by someone or other (i.e., $\mathbf{T}^{\text{Ind}} >_{\mathbf{V}} \mathbf{T}^{\text{Non}}$), the extra richness of the former may not fully make up for the difference in richness between \mathbf{W}^{Dir} and \mathbf{T}^{Non} . Hence, directly participating in the preservation of the Wetlands may still be a richer property to have than indirectly participating in the Takahe preservation.

So the following forms a consistent set of value relations:

$$\mathbf{W}^{\text{Dir}} >_{\mathbf{V}} \mathbf{T}^{\text{Ind}} >_{\mathbf{V}} \mathbf{T}^{\text{Non}} >_{\mathbf{V}} \mathbf{W}^{\text{Non}}.$$

Assuming these are the actual value relations among these properties—it's my story!—it would be entirely fitting for Erin's preferences to be isomorphic to these, and we can endorse that judgement without having to invoke the notions of value distance and value perspective that we need to invoke when we take states to be value bearers.

Note that Tammy shares some of these preferences. She prefers \mathbf{T}^{Ind} to \mathbf{T}^{Non} and prefers \mathbf{T}^{Non} to \mathbf{W}^{Non} . But Tammy is not at all desirous of giving up *her* lucrative job and participating directly in the preservation of the Wetlands. This does not generate a discrepancy between her and Erin, however. Since Tammy does not bear any pre-existing rich relations to the Wetlands, she is not eligible to possess \mathbf{W}^{Dir} . So her decision to devote her money to the Takahe preservation fund does not show that she prefers \mathbf{T}^{Ind} to \mathbf{W}^{Dir} . Even if she gave up her job in real estate, she could not acquire the property \mathbf{W}^{Dir} by going to work with Erin, since she would simply not have the prerequisites for it.

But what about the fact that Tammy prefers the *indirect* preservation of the Takahe (i.e., by keeping her job and donating the cash) to direct preservation (i.e., by giving up her job and working for the Takahe Foundation)? Is this not a case where her preferences, by failing to align with the real values, are defective? It is possible, but it may be also be that these are not the salient properties for Tammy to be weighing. It may be that she simply cannot possess \mathbf{T}^{Dir} . Recall that Tammy has no aptitude for that kind of work, and so any attempt on her part

to acquire T^{Dir} will end in complete failure. Or it might also be that although Tammy really does prefer T^{Dir} to T^{Ind} , her choice is really T^{Dir} and some property P that is logically stronger and richer than T^{Ind} , a property which implies T^{Ind} , excludes T^{Dir} , and which is much richer than either of those. It might be that by staying on and making loads of cash in real estate she can instantiate many more rich properties (she might, for example, have the opportunity to learn to play the *Goldberg Variations*).

The same sorts of considerations might apply to Erin's preference for W^{Dir} over T^{Dir} . She bears close causal, historical, and epistemic connections to the Wetlands ecosystem by virtue of her many years of interacting with them. By contrast, she bears none of these connections to the Takahe ecosystem, so she cannot acquire T^{Dir} even if she wanted to. Further, even if the Takahe's being preserved, T^{Non} , is a richer global property than the Wetlands' being preserved, W^{Non} , directly preserving the Wetlands, W^{Dir} , is a richer local property than the Takahe's being preserved (T^{Non}) or even indirectly preserving the Takahe (T^{Ind}).

Finally, all of this is compatible with the assumption that the global property of the Takahe's being preserved appears more valuable to Erin than does the Wetlands' being preserved. So Erin can fittingly prefer the local property of directly preserving the Wetlands to the local property of indirectly preserving the Takahe, even though she prefers the state consisting in the Takahe ecosystem's being preserved to the state consisting in the Wetlands ecosystem's being preserved.

The property account of value bearers gives the fitting attitude theorist a wide scope for accounting for apparently differing preferences among valuers without deeming some of them defective and without having to appeal to value distance or value perspective. The work done by value distance and value perspective on the state theory is accomplished instead by the differing degrees of richness of the connections that a valuer can bear to those independently rich systems.

This leaves one last problem. Why should desires line up with the appearances of goodness/richness in the way that the value appearance thesis says they should? This is a particularly stark problem if states of affairs are the basic bearers of value. Suppose you are contemplating two states of affairs and it becomes clear to you that one of them has a slightly higher degree of unified complexity than the other. Why should the former necessarily strike you as *better*? Why should you *ipso facto* prefer the former to the latter?

Suppose an asteroid is heading toward a pair of twin inhabited planets, and it is not clear which one it is going to hit and destroy. The planet you are on (call it *Planet*) has slightly smaller degree of unified complexity than its twin does (call that one *Twin Planet*). When you are huddling with your family waiting for calamity to strike either Planet or Twin Planet, should you prefer that Planet, along with your nearest and dearest, be obliterated lest the universe suffer a slightly greater diminution of overall richness?

Now suppose that properties, rather than states, bear value. If the asteroid hits your planet, most of the properties you will end up with will be very impoverished indeed. You, and all the folk with whom you enjoy rich connections, will be reduced to rubble. That is an extremely poor set of properties to end up instantiating. But if it is fitting for you to prefer richer properties to poorer properties, then it is fitting for you to prefer surviving largely intact to being obliterated, even if this results in some very small overall cost to the overall richness of the world. It is better to be rich than to be poor, and so it is better to survive with a high degree of richness than to be reduced to a heap of rubble¹⁹.

The combination of these three theses—of properties as value bearers, of value as richness, and of preference as the fitting response to betterness—yields the following important consequence: it is entirely fitting for you to prefer the richer local property of *surviving intact along with your nearest and dearest while the other planet is destroyed* (S) to the poorer local property of *being reduced to rubble along with your nearest and dearest while the other planet survives* (R). And your twin on Twin Planet can, entirely fittingly, also prefer S to R. Thus, your preferences can be perfectly isomorphic both to the value relations themselves as well as to each other.

But here is a possible objection. The global property of *Twin Planet's surviving while Planet is reduced to rubble* (TS) is (by assumption) a slightly richer property than the global property of *Planet's surviving while Twin Planet is reduced to rubble* (PS). So it would also be fitting for you to prefer global property TS to global property PS. But then we seem to have an irresolvable conflict between fitting preferences over global properties and fitting preferences over local properties. (There is no such conflict for your more fortunately placed twin.) Possessing the richer of the two local properties, in these circumstances, precludes possessing the richer of the two global properties. Your preferences over local properties and over global properties, despite both being fitting according to the theory, thus come into irresolvable conflict. If your preferences over global properties are eventually satisfied (that is, Planet is reduced to rubble while Twin Planet survives) your preferences over the local properties are not (you, along with your nearest and dearest, become a bloody scattered mess in the rubble, an extremely poor state to end up in) and vice versa. Since not all your fitting preferences can be satisfied, this combination of theses is apparently incoherent.

While much more needs to be said about this objection, I don't think it delivers a fatal blow to the combination. Fitting preferences should track the values. But for almost any theory of value, there will be occasions when not all values can be realized, and so not all fitting preferences can be satisfied either. In this case there are just two ways of combining local value (local richness) with global value (global richness). One possibility combines the realization of vastly greater local richness, but at a tiny cost to global richness; and the other possibility combines an enormous diminution to local richness with a tiny gain in global richness. With that way of framing it, it does not seem totally irrational to prefer the former to the latter, all things considered.

CONCLUDING REMARKS

The Isomorphic Response objection suggests that at least one of our two environmental heroines—Erin or Tammy—is responding unfittingly because their preferences appear not to be isomorphic to each other. Our three strategies all give the fitting attitude theorist who embraces value as richness some resources for resisting the objection and deeming both Tammy and Erin’s responses fitting despite their apparent differences.

The first strategy countenances a range of permissible preferences at the outset. While the betterness relation serves to constrain the fittingness of responses, it does not prescribe one unique preference ordering to all rational valuers. There can thus be multiple different yet legitimate responses to richness, and two fitting responders can have non-isomorphic preferences. But even though this gives responders some latitude in their preferences, it does not permit any responder to prefer one state of affairs to another state of affairs that is equally rich. So it is unfitting for Erin to prefer the preservation of her beloved Wetlands to the preservation of some distant but equally rich ecosystem with which she has no connection. Since these are equally rich ecosystems, she should simply be indifferent between them. And if Erin *thinks* that they are equally rich, then it would be incoherent of her not to be indifferent.

The second strategy takes the notion of value appearances seriously and attributes a perspectival aspect to such appearances. Perceptual representation depends not only on the properties of the perceived objects themselves but on the relation of the perceiver to those objects. Objects that are closer to a perceiver should loom larger for her in her visual perception than objects that are more distant. Analogously, some loci in value space are closer to a responder than are others; accordingly, it is sometimes fitting for one of two equally rich states to *appear* better than the other to a particular responder. So it is fitting for the preservation of the Wetlands to *seem* better to Erin than the preservation of the distant Grasslands. If preferences are value appearances, then it is fitting for Erin to *prefer* the preservation of the Wetlands to the preservation of the Grasslands: even if she *thinks* that the Grasslands are no less rich than her Wetlands, she can nevertheless legitimately ignore the Grasslands and instead work assiduously to preserve her beloved Wetlands.

The third strategy takes properties, rather than states of affairs, to be value bearers. Global properties (those that are had by all if had by one) can do double-duty as states. But there are many local properties that are both objects of desire as well as bearers of value. The value of a property is determined by its richness, just as (on the value as richness account) the value of a state is determined by its richness. For Erin, preserving her beloved Wetlands is a richer property to have than the property of preserving the remote and unloved Grasslands. This is quite compatible with the Wetlands’ being no richer than the Grasslands. Some relations which Erin can bear to the Wetlands are rich not solely in virtue of the complexity of the Wetlands themselves, but also in virtue of her connections to the

Wetlands. And since she bears these relations to the Wetlands but not to the Grasslands, she will become richer by preserving the former rather than the latter. Her preference for one local property over the other is not only fitting (the former is richer than the latter) but is also compatible with her thinking the two ecosystems to be equally rich and equally valuable in themselves.

The fitting attitude theorist is not necessarily forced to choose just one of these strategies. The first and the third seem to be compatible. One might hold that there is more than one permissible preference ordering on the grounds that richness does not induce a single complete ordering of properties. Further, one could combine this position with the idea that some properties are determinately richer by virtue of relations between the property-bearer and other entities that the property involves. One might also combine the second and the third strategies. It may be that directly preserving the Wetlands is preferable to indirectly preserving the Takahe ecosystem due to the richness of the respective properties. But it might also be that even if the former is not a richer property, it should loom larger in Erin's desires simply because the Wetlands are closer to her in value space. It does not seem possible, however, to combine the first and second strategies. For the first requires that Erin be indifferent between equally good properties, whereas the second may require her to prefer one to the other.

What is clear, I think, is that the combination of value as richness, a fitting attitudes account of the representational kind, and properties as value bearers yields a rich array of resources for environmental value theory that deserve further exploration²⁰.

NOTES

- ¹ See Kelly 2014 (this volume) and Kelly 2004 for an exposition of the idea of value as richness and its own rich history. The idea can be found explicitly in both Leibniz and Nozick, but it is also implicit in a lot of work on value theory. One can take the proposal, as Kelly does, to constitute a complete and monistic theory of value, in which goodness is in some sense identified with richness. Or, perhaps more plausibly, one can take it to be a partial account of value, in which richness is one of a number of thick determinates of the thin concept of goodness. For my part, I am inclined to think that the latter is the more plausible, but for the purposes of this paper I can leave this question open.
- ² Rabinowicz 2013.
- ³ Not all fitting attitude enthusiasts are value reductionists by any means. See, for example, Zimmerman 2001 and Wedgwood 2009. The latter argues explicitly that it is a mistake to conflate the universal applicability of the FA biconditional (if it is indeed universal) with the idea that the normative concept of fittingness has some kind conceptual priority with respect to the axiological concepts (p. 516). Further, as I argue below, one can embrace the universal applicability of FA biconditionals while adopting a *non-normative* notion of fittingness—fittingness as representational correctness—and this combination clearly eschews the possibility of reduction altogether. So one can embrace the FA biconditional schema without embracing the idea that the FA schema yields a reductive analysis of value.
- ⁴ Rabinowicz and Rønnow-Rasmussen 2004.
- ⁵ Kelly 2004 and 2014.
- ⁶ See Kelly 2014 : “God has chosen the [rule] which is the most perfect, that is to say the one which is at the same time the simplest in hypotheses and the richest in phenomena, as a geometrical line might be, of which the construction was easy and the properties and effects very admirable and of great extent.” Leibniz 1953, p. 9. “Holding fixed the unifiedness of the material, the degree of organic unity varies directly with the degree of diversity of that material being unified. Holding fixed the degree of diversity of the material, the degree of organic unity varies directly with the degree of unifiedness [...]” Nozick 1981, p. 416.
- ⁷ The Isomorphic Response objection was raised in Oddie 2005. It was also raised by A.C. Ewing as an objection to his own fitting attitudes account. See Olson 2009, pp. 365–378.
- ⁸ Rabinowicz 2008.
- ⁹ See the quote from Nozick 1981, footnote vi.
- ¹⁰ Rabinowicz and Rønnow-Rasmussen 2004.
- ¹¹ Tappolet 2011, p. 119.
- ¹² St. Augustine 1982, p. 109.
- ¹³ For a longer development of the notions of distance in value space and of its perspectival effects, see Oddie 2005, chapter 8.
- ¹⁴ See Zimmerman 2010, section 4.
- ¹⁵ Lewis 1979. Note that one does not have to buy into the baroque aspects of Lewisian metaphysics to find the property view of desire plausible. Indeed, the view is probably more plausible if worlds are abstract ways that things can be, rather than large, causally isolated hunks of junk. Nor does one have to buy into Lewis’s companion thesis that the objects of belief are also properties, though that view sits happily enough with what I am endorsing here.
- ¹⁶ This is not how Lewis characterizes the property view in his 1979 paper, but his characterization is dependent on his idiosyncratic view of individuals and worlds—in particular his denial of any trans-world identity.
- ¹⁷ This talk of depth of specification is not just a metaphor. The depth of a characterization is a measure of the number of nested quantifiers that one needs to use to obtain that characterization. What I have been describing informally here are the depth-*d* distributive normal forms discovered by Hintikka (see his 1963 paper). These have been put to good philosophical work in both inductive logic and truthlikeness. See Oddie 1986 chapter 4 for an accessible introduction.

- ¹⁸ As noted by an anonymous referee, other things may not always be equal. If Bach listened to some amateur playing the *Goldberg Variations*, the relation of Bach's mental state to the played music would doubtless add far greater richness to the world than that added by the relation of the amateur's mental state to the played music.
- ¹⁹ Incidentally, this example also gives us a glimpse of how VR might be combined with the thesis that properties are value bearers to defuse the evolutionary debunking argument against value realism. But that is a topic for another paper.
- ²⁰ I would very like to thank the sponsors and organizers of the one-day workshop on *Attitudes, Values, Environment* at the University of Montreal in February 2014 (particularly Christine Tappolet), and the two anonymous referees for this journal, for many excellent comments and criticisms on earlier versions of this paper.

REFERENCES

- St. Augustine, "The Literal Meaning of Genesis," in J. Quasten, W.J. Burghardt and T.C. Lawler (eds.), *Ancient Christian Writers: Vol. 1 Book 4*, New York/Mahwah, NJ, Paulist Press, 1982.
- Hintikka, Jaakko, "Distributive normal forms in first-order logic," in J.N. Crossley and M.A.E. Dummett (eds.), *Formal Systems and Recursive Functions (Proceedings of the Eighth Logic Colloquium)*, Amsterdam, North-Holland, 1963, pp. 47-90.
- Kelly, Crhis, *A Theory of the Good*, Ph.D. Thesis, University of Colorado, Boulder, CO, 2003.
- Kelly, Chris, "Value Monism, Richness, and Environmental Ethics," *Les ateliers de l'éthique/The Ethics Forum*, vol. 9, no. 2, 2014, pp. 100-129.
- Leibniz, Gottfried Wilhelm, *Discourse on Metaphysics*, Manchester, Manchester University Press, 1953.
- Lewis, David, "Attitudes *De Dicto* and *De Se*," *The Philosophical Review*, vol. 88, no. 4, 1979, pp. 513-543.
- Nozick, Robert, *Philosophical Explanations*, Cambridge, MA, Harvard University Press, 1981.
- Oddie, Graham, *Likeness to Truth*, Dordrecht, Reidel, 1986.
- Oddie, Graham, *Value, Reality and Desire*, Oxford, Oxford University Press, 2005.
- Olson, Jonas, "Fitting Attitude Analyses of Value and the Partiality Challenge," *Ethical Theory and Moral Practice*, vol. 12, no. 4, 2009, pp. 365-378.
- Rabinowicz, Wlodek, "Value relations," *Theoria*, vol. 74, no. 1, 2008, pp. 18-49.
- Rabinowicz, Wlodek, "Value, Fitting Attitude Account of," in Hugh LaFollette (ed.), *The International Encyclopedia of Ethics*, Oxford, Wiley-Blackwell, 2013.
- Rabinowicz, Wlodek and Toni Rønnow-Rasmussen, "The Strike of the Demon: on Fitting Pro-attitudes and Value," *Ethics*, vol. 114, no. 3, 2004, pp. 391-423.
- Tappolet, Christine, "Values and Emotions: Neo-Sentimentalism's Prospects," in Carla Bagnoli (ed.), *Morality and the Emotions*, Oxford, Oxford University Press, 2011.
- Wedgwood, Ralph, "The "Good" and the "Right" Revisited," *Philosophical Perspectives*, vol. 23, no. 1, 2009, pp. 499-519.
- Zimmerman, Michael J., *The Nature of Intrinsic Value*, Lanham, Rowman and Littlefield, 2001.
- Zimmerman, Michael J., "Intrinsic vs. Extrinsic Value," Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2010 Edition, URL = <<http://plato.stanford.edu/archives/win2010/entries/value-intrinsic-extrinsic/>>.

SUR LA SYMÉTRIE PRÉSUMÉE ENTRE VALEURS ET PRÉFÉRENCES¹

MAURO ROSSI
UNIVERSITÉ DU QUÉBEC À MONTRÉAL

RÉSUMÉ :

Comment pouvons-nous analyser des relations de valeur non standards, comme la parité axiologique, en termes d'attitudes appropriées? Wlodek Rabinowicz suggère que deux choses sont à parité si et seulement si il est à la fois permissible de préférer l'une à l'autre et permissible d'avoir la préférence contraire. Dans un article récent, Johan Gustafsson soutient toutefois que l'analyse de Rabinowicz viole un principe de symétrie entre valeurs et préférences, selon lequel il existe pour toute relation de valeur une relation de préférence correspondante (et vice-versa). À la lumière de ce principe, Gustafsson propose une analyse alternative, selon laquelle deux choses sont à parité si et seulement si il est requis d'entretenir ces choses en parité préférentielle. Dans cet article, j'examine en détail les arguments avancés par Gustafsson contre l'analyse de Rabinowicz et je montre qu'aucun d'eux n'est convaincant.

ABSTRACT :

How can we analyse non-standard value relations, such as axiological parity, in terms of fitting attitudes? Wlodek Rabinowicz has suggested that two items are on a par if and only if it is both permissible to prefer one to the other and permissible to have the opposite preference. In a recent paper, however, Johan Gustafsson has argued that Rabinowicz's account violates a principle of value-preference symmetry, according to which for any value relation, there is a corresponding preference relation (and vice-versa). Following this principle, Gustafsson has proposed an alternative analysis, according to which two items are on a par if and only if one is required to hold these items preferentially on a par. In this paper, I examine in detail Gustafsson's arguments against Rabinowicz's analysis and show that none of them succeeds.

INTRODUCTION

Nous effectuons souvent des comparaisons concernant la valeur de différentes choses. En général, nous jugeons qu'une chose est mieux que, égale à, ou pire qu'une autre. Parfois, nous jugeons, toutefois, que la valeur de deux choses est incomparable. Il s'agit là de ce que je nommerai des relations de valeur *standards*. Selon certains philosophes, ces relations n'épuisent pas l'espace conceptuel des relations de valeur. En effet, certains pensent qu'il existe une quatrième relation de valeur positive, à savoir la parité, qui vaut entre deux choses lorsqu'aucune des relations de valeur standards ne s'applique (Chang, 1997, 2002b). Je nommerai relations de valeur *non standards* les relations de valeur comme la parité.

Intuitivement, des relations de valeur non standards comme la parité pourraient valoir dans des cas où la comparaison de deux choses par rapport à une valeur de référence dépend de plusieurs dimensions différentes. Supposons, par exemple, que l'on compare le joueur de tennis Rafael Nadal et le joueur de football Leo Messi du point de vue de leur talent sportif. Il est plausible de penser que le talent sportif est déterminé par plusieurs facteurs, tels que la créativité, la technique individuelle, le caractère, la capacité de « lire » le match, et ainsi de suite. Or, *si* aucun des deux joueurs n'est supérieur, ou égal, à l'autre dans chacune de ces dimensions, et *si* ces dimensions peuvent être pesées de manières différentes, mais également justifiées, nous pourrions conclure que Nadal et Messi ne sont liés par aucune des relations de valeur standards, mais qu'ils sont plutôt à parité (*on a par*) eu égard à leur talent sportif.

La question de savoir si cette conclusion est justifiée et, plus généralement, celle de savoir si la relation de parité axiologique a de réelles applications a suscité beaucoup d'intérêt dans la littérature (Chang, 1997, 2002b; Hsieh, 2005; Peterson, 2007; Espinoza, 2008; Gustafsson et Espinoza, 2010; Carlson, 2011; Gustafsson, 2013b). Toutefois, dans cet article, je m'intéresserai à une question différente : celle de savoir comment nous pouvons *analyser* les relations de valeur standards et non standards. Plus spécifiquement, je considérerai les questions suivantes : qu'est-ce que cela signifie de dire que deux choses sont à parité axiologique? Qu'est-ce que cela signifie de dire qu'elles sont liées par une des autres relations de valeur?

Une stratégie intuitive pour répondre à ces questions consiste à considérer les analyses traditionnelles de la valeur *non comparative* et à les appliquer aux relations de valeur. Au cours des dernières années, plusieurs auteurs se sont tournés vers une analyse de la valeur en termes d'attitudes appropriées (*fitting-attitude analysis*) (Scanlon, 1998; Rabinowicz et Rønnow-Rasmussen, 2004; Danielsson et Olson, 2007). En ce qui concerne le cas non comparatif, l'analyse de la valeur en termes d'attitudes appropriées soutient qu'une chose est bonne si et seulement si elle est l'objet approprié d'une attitude favorable. Le concept de bien (*goodness*) est ainsi analysé en fonction d'une composante normative, comprise en termes de la notion de « approprié » (*fittingness*), et d'une composante

psychologique, comprise en termes de la notion de « attitude favorable » (*favouring*). À supposer que cette analyse offre une conception correcte de la valeur non comparative, comment pouvons-nous alors l'élargir de telle sorte qu'elle puisse offrir une caractérisation des relations de valeur standards et non standards?

Dans une série d'articles récents, Rabinowicz propose une analyse de la valeur comparative fondée sur trois caractéristiques principales (Rabinowicz, 2008; 2009; 2012)². La première consiste en ce que nous pourrions appeler la thèse générique de l'analyse de la valeur comparative en termes d'attitudes appropriées. Elle stipule que tout jugement de valeur comparatif équivaut à une évaluation normative des préférences. La seconde consiste en une thèse au sujet de la composante normative de l'analyse, affirmant que la normativité est caractérisée par deux niveaux distincts, à savoir le niveau du « permmissible » et celui du « requis » (*required*). La troisième consiste en une thèse au sujet de la composante psychologique de l'analyse, affirmant que les seules relations de préférence qui existent sont les relations de préférence standards, à savoir la préférence stricte, l'indifférence, et – Rabinowicz ajoute – la relation d'absence de préférence (*preferential gap*). L'idée de Rabinowicz est que nous pouvons esquisser une analyse des relations de valeur standards et non standards à partir des relations de préférence standards et des deux niveaux de normativité mentionnés.

Il semble évident que le cadre théorique de Rabinowicz incorpore plus de relations de valeur que de relations de préférence. Pourtant, cet aspect de son analyse n'est pas parvenue à convaincre tout le monde. Dans un article récent, Johan Gustafsson fait valoir que les relations de valeur et les relations de préférence sont symétriques (Gustafsson, 2013a). Selon lui, il existe pour toute relation de valeur une relation de préférence correspondante, et vice-versa. C'est ce que Gustafsson nomme le *principe de symétrie des valeurs et des préférences*. Le cadre théorique de Rabinowicz semble aller à l'encontre de ce principe puisqu'il ne recourt pas aux relations de préférence non standards afin de rendre compte des relations de valeur non standards³. Gustafsson suggère de substituer à l'analyse de Rabinowicz une tout autre analyse en termes d'attitudes appropriées, en vertu de laquelle toutes les relations de valeur (standards et non standards) sont analysées en fonction des relations de préférence (standards et non standards) correspondantes et d'un seul niveau de normativité : celui du « requis »⁴.

L'idée de Gustafsson est que, si le principe de symétrie entre les valeurs et les préférences est valide, alors, *ceteris paribus*, son analyse devrait être préférée à celle de Rabinowicz. Comme nous le constaterons, l'analyse de Rabinowicz associe à toute relation de valeur standard une relation de préférence standard. Ce n'est que lorsqu'il est question de relations de valeur non standards que la symétrie entre les valeurs et les préférences cesse de prévaloir. Toutefois, si le principe de symétrie des valeurs et des préférences est valide, il semble s'ensuire, selon Gustafsson, que, par souci de simplicité et de cohérence interne, la symétrie de l'analyse doit également être maintenue eu égard aux cas de relations

de valeur non standards. D'autre part, ces considérations appuient l'analyse de Gustafsson seulement si toutes choses demeurent égales par ailleurs, c'est-à-dire seulement si l'analyse de Gustafsson est à même d'expliquer tous les phénomènes dont rend compte l'analyse de Rabinowicz. D'où la nécessité d'imposer une clause *ceteris paribus*.

Gustafsson présente deux arguments en faveur de son analyse de la valeur comparative en termes d'attitudes appropriées. Le premier a pour but de montrer que le principe de symétrie des valeurs et des préférences est valable. Le second a pour but de montrer que la clause *ceteris paribus* est satisfaite. Plus précisément, le premier argument de Gustafsson fait valoir que toutes les raisons que l'on peut invoquer pour et contre l'existence de relations de valeur non standards constituent également des raisons probantes pour et contre l'existence de relations de préférence non standards. Par conséquent, s'il y a des relations de valeur non standards, il existe également des relations de préférence non standards correspondantes, comme le maintient le principe de symétrie des valeurs et des préférences. Le second argument présenté par Gustafsson fait valoir que sa théorie peut accommoder l'idée qui motive à l'origine la théorie de Rabinowicz, c'est-à-dire l'idée que, dans certains cas, il est également permmissible d'entretenir des relations de préférence différentes à l'endroit d'une paire de choses.

L'objectif de cet article est de montrer que les arguments proposés par Gustafsson contre l'analyse de la valeur en termes d'attitudes appropriées de Rabinowicz ne parviennent pas à leur fin. Avant d'examiner ces arguments, je présenterai plus en détail les conceptions respectives de Rabinowicz et de Gustafsson concernant l'analyse en termes d'attitudes appropriées de la valeur comparative.

RABINOWICZ ET GUSTAFSSON SUR L'ANALYSE EN TERMES D'ATTITUDES APPROPRIÉES DES RELATIONS DE VALEUR

Tel qu'indiqué dans l'introduction, l'analyse en termes d'attitudes appropriées des relations de valeur proposée par Rabinowicz repose sur une distinction entre deux niveaux de normativité, c'est-à-dire le « permmissible » et le « requis ». En fait, l'analyse des relations de valeur standards proposée par Rabinowicz implique seulement le niveau du « requis ». Selon Rabinowicz, pour toute option x et pour toute option y :

- (B) x est *mieux* qu' y si et seulement si il est requis de préférer x à y .
- (W) x est *pire* qu' y si et seulement si il est requis de préférer y à x .
- (E) x est *aussi bon* qu' y si et seulement si il est requis d'être indifférent à l'endroit de x et d' y .
- (I) x est *incomparable* à y sur le plan de la valeur si et seulement si il est requis d'entretenir une absence de préférence entre x et y .

Le second niveau de normativité, celui de la permmissibilité, entre en jeu seulement lorsqu'il s'agit de caractériser les relations de valeur non standards. Considérons la parité. Comme nous avons pu le constater, deux choses sont à parité (*on a par*) lorsqu'aucune des deux n'est ni mieux que l'autre, ni également

bonne, alors que les deux choses s'avèrent pourtant comparables sur le plan de la valeur. Dans le cadre de l'analyse de la valeur en termes d'attitudes appropriées proposée par Rabinowicz, la parité est caractérisée comme suit :

(P_R) x est à parité avec y si et seulement si il est à la fois permissible de préférer x à y et permissible de préférer y à x .

Rabinowicz offre une interprétation plus formelle de cette analyse de la valeur en termes d'attitudes appropriées d'après l'intersection des classements de préférence permisibles. Sur la base de ce cadre théorique, Rabinowicz est à même de distinguer quinze relations de valeur atomiques. Pour m'en tenir à mon objectif actuel, il suffira de considérer seulement l'analyse informelle de Rabinowicz et de prendre la parité comme exemple paradigmatique de relation de valeur non standard.

Contrairement à Rabinowicz, Gustafsson pense qu'il est possible d'analyser les relations de valeur standards et non standards sur la base d'un seul niveau de normativité, celui du « requis ». L'analyse des relations de valeur standards qu'il propose est identique à celle de Rabinowicz. La différence entre les deux théories réside dans l'analyse des relations de valeur non standards. Considérons la parité. Afin d'analyser le concept de parité au sein d'un cadre théorique qui n'inclut qu'un seul niveau de normativité, Gustafsson propose d'élargir l'ensemble des relations de préférence et d'accommoder une relation de préférence non standard additionnelle, à savoir la relation de parité préférentielle. Il suggère l'analyse suivante :

(P_G) x est à parité avec y si et seulement si il est requis d'entretenir x et y en parité préférentielle.

Cette stratégie peut se généraliser. Selon Gustafsson, pour toute relation de valeur (standard et non standard), il existe une relation de préférence correspondante (standard et non standard) en fonction de laquelle toute relation de valeur peut être analysée. Gustafsson soutient que cela nous offre une analyse plus plausible de la valeur comparative en termes d'attitudes appropriées.

LE PREMIER ARGUMENT DE GUSTAFSSON

Le point de départ du premier argument de Gustafsson consiste à faire remarquer qu'il existe un rapport de correspondance entre les relations de valeur et de préférence lorsque nous considérons les relations standards. Autrement dit, lorsqu'il est question de relations standards, valeurs et préférences sont symétriques. Gustafsson affirme que si la symétrie se décompose, cela doit être parce qu'il existe une relation de valeur non standard à laquelle ne correspond aucune relation de préférence spécifique. Afin de montrer que cela est effectivement le cas, on doit recourir à un argument convaincant pour justifier l'existence d'une relation de valeur non standard, qui n'est pas, *mutatis mutandis*, un argument convaincant pour justifier l'existence d'une relation de préférence non standard. C'est ici que le problème apparaît, selon Gustafsson, étant donné que tous les arguments pour

ou contre l'existence de relations de valeur non standards semblent aussi bien constituer des arguments pour et contre l'existence de relations de préférence non standards. Or si cela est effectivement le cas, nous devrions être en mesure de conclure que les relations de valeur et les relations de préférence sont *toujours* symétriques, qu'elles soient standards ou non standards. Étant donné que l'analyse de la valeur comparative en termes d'attitudes appropriées de Rabinowicz est incompatible avec une telle conclusion, nous aurions raison alors de préférer l'analyse de Gustafsson à celle de Rabinowicz.

Quels sont les arguments typiques avancés en faveur des relations de valeur non standards? Les deux arguments principaux discutés dans la littérature sont l'argument de la faible amélioration (*small improvement argument*) et l'argument de l'enchaînement (*chaining argument*) (Chang, 2002b). Le premier a pour but de montrer qu'il existe des cas où pour certains x et y , x n'est pas mieux qu' y , y n'est pas mieux que x , et x n'est pas aussi bon que y . Le second argument consiste à faire valoir que ces cas ne sont pas des cas d'incomparabilité axiologique, c'est-à-dire que x et y ne sont pas incomparables sur le plan de la valeur. Il s'agit plutôt de cas où une relation de valeur additionnelle intervient, à savoir la parité.

Gustafsson pense que ces arguments peuvent également être appliqués aux relations de préférence (requis). Lorsqu'ils sont adaptés en conséquence, l'argument de la faible amélioration et l'argument de l'enchaînement sont censés montrer que, dans certains cas, il n'est ni strictement requis de préférer x à y , ni strictement requis d'être indifférent à l'endroit de x et de y , ni strictement requis de préférer y à x , alors que, pourtant, il est requis de ne pas entretenir une relation d'absence de préférence entre x et y .

Il convient de noter que la possibilité de procéder à une reformulation de ces arguments en termes de relations de préférence requises n'est pas tout à fait surprenante. En fait, cela découle directement de l'analyse en termes d'attitudes appropriées des relations de valeur standards, à laquelle Rabinowicz et Gustafsson souscrivent tous deux. Selon Gustafsson, la possibilité d'une telle reformulation soulève, toutefois, une question importante. Si, d'un point de vue axiologique, la combinaison de l'argument de la faible amélioration et de l'argument de l'enchaînement suffit pour conclure qu'il existe une quatrième relation de valeur positive, à savoir la parité axiologique, pourquoi ne pourrions-nous pas également conclure, dans le cas des préférences, qu'il existe une quatrième relation de préférence positive, à savoir la parité préférentielle, que nous sommes requis d'avoir? Si l'argument de la faible amélioration et l'argument de l'enchaînement s'appliquent au cas préférentiel, nous avons alors des raisons de croire que les relations de valeur et les relations de préférence sont symétriques⁵.

Dans la même veine, Gustafsson pense que les objections adressées à l'argument de la faible amélioration et à l'argument de l'enchaînement peuvent être reformulées de façon à ce qu'elles valent pour les relations de préférence (requis). Dans le cas axiologique, ces objections font valoir que la parité axio-

logique est, au mieux, une relation dérivée des relations de valeur standards. Dès lors, la question suivante s'impose : si de tels arguments sont jugés convaincants en ce qui concerne le cas axiologique, pourquoi n'en irait-il pas de même pour le cas préférentiel? Plus précisément, pourquoi ces arguments ne peuvent-ils pas démontrer que la relation de parité préférentielle est, au mieux, une relation dérivée des relations de préférence standards? Autrement dit, si les objections contre l'argument de la faible amélioration et l'argument de l'enchaînement valent également pour le cas préférentiel, il semble qu'il n'y ait aucune raison de penser que les relations de valeur et les relations de préférence sont asymétriques.

Par ailleurs, il convient de souligner que Gustafsson ne prend pas parti dans ce débat sur la parité. Ce qui importe pour lui, c'est justement que les deux arguments avancés en faveur de l'existence de la parité en tant que relation non dérivée, tout comme les objections qui leur sont adressées, valent aussi bien pour le cas préférentiel que pour le cas axiologique. En fait, l'argument de Gustafsson en faveur du principe de symétrie entre les valeurs et les préférences est *indépendant* de tout ce qui peut résulter du débat substantiel sur la parité. Son argument consiste simplement à faire valoir que s'il existe effectivement une relation de parité axiologique (dérivée ou non dérivée), il doit également exister une relation de parité préférentielle (dérivée ou non dérivée), et vice-versa.

DISCUSSION DU PREMIER ARGUMENT DE GUSTAFSSON

À première vue, l'argument de Gustafsson semble convaincant. Dans cette section, je soutiendrai toutefois que son argument est plus problématique qu'il ne le paraît. Plus précisément, je soutiendrai, d'une part, que la validité de l'argument de Gustafsson n'est pas indépendante de la caractérisation des préférences que l'on adopte et, d'autre part, que nous avons raison de nous intéresser à la façon dont les préférences sont caractérisées.

Dans son article, Gustafsson ne précise pas quelle conception des préférences il considère, et il ne croit pas qu'il s'agit là d'une question importante. Il mentionne seulement que si les préférences sont des jugements de valeur, le principe de symétrie entre les valeurs et les préférences s'ensuit alors de manière évidente⁶. Si les préférences sont des jugements de valeur, entretenir une relation de préférence entre deux choses équivaut à juger qu'il y a une relation de valeur correspondante entre ces choses⁷. Par exemple, préférer x à y consiste à juger que x est mieux qu' y ; être indifférent entre x et y consiste à juger que x est aussi bon qu' y ; entretenir une relation d'absence de préférence entre x et y consiste à juger que x et y sont incomparables sur le plan de la valeur. La parité préférentielle peut être, dès lors, caractérisée ainsi : entretenir deux choses, x et y , en parité préférentielle consiste simplement à juger que x est axiologiquement à parité avec y . Cette caractérisation appuie manifestement la position de Gustafsson. En effet, si la parité axiologique existe, la parité préférentielle existe également, car la parité préférentielle consiste simplement dans le jugement que la parité axiologique vaut entre deux choses.

Toutefois, une telle caractérisation ne garantit pas à Gustafsson une victoire facile. En fait, il y a de bonnes raisons de croire que la conception des préférences en tant que jugements de valeur ne peut jouer le rôle qui lui revient dans le cadre de l'analyse de la valeur comparative en termes d'attitudes appropriées. Cela est aisé à comprendre à la lumière de l'analyse de la parité axiologique qui découle de cette conception :

(P_G') x est axiologiquement à parité avec y si et seulement si il est requis de juger que x et y sont axiologiquement à parité.

Cet énoncé semble vrai. Toutefois, ce qu'il est important de noter est que (P_G') n'offre pas une analyse *réductive* du concept de parité axiologique. En effet, le concept de parité axiologique intervient à la fois du côté gauche et du côté droit de (P_G') . Par conséquent, si les préférences sont conçues comme des jugements de valeur, il est impossible de formuler une analyse *réductive* des relations de valeur en termes de relations de préférence. On peut toutefois affirmer que cela est précisément le but de l'analyse de Rabinowicz. Ce qui en résulte est qu'une conception des préférences en termes de jugements de valeur ne peut jouer le rôle qui devrait lui incomber dans le cadre de l'analyse des relations de valeur en termes d'attitudes appropriées⁸.

Supposons maintenant que nous considérons une conception des préférences qui convient davantage à cet objectif. Considérons, par exemple, la conception des préférences en tant que relations entre attitudes favorables monadiques, telle qu'elle est développée par Rabinowicz⁹. Selon cette conception, préférer strictement x à y équivaut à favoriser x davantage qu' y ; être indifférent entre x et y équivaut à favoriser x au même titre qu' y ; entretenir une relation d'absence de préférence entre x et y équivaut à favoriser x et y avec des degrés incommensurables. La parité préférentielle peut alors être caractérisée ainsi : x est *préférentiellement à parité* avec y si et seulement si x et y sont favorisés avec des degrés qui sont à parité. De là, nous obtenons l'analyse suivante de la parité axiologique.

(P_G'') x est axiologiquement en parité avec y si et seulement si il est requis de favoriser x et y à des degrés qui sont à parité.

(P_G'') offre une analyse réductive de la parité axiologique. Si nous adhérons ainsi à une conception des préférences en termes de relations entre attitudes monadiques, nous ne sommes pas confrontés à un problème de circularité tel qu'il est soulevé par la conception selon laquelle les préférences sont des jugements de valeur. Toutefois, nous sommes alors confrontés à un problème d'un tout autre ordre : ce que signifie l'idée selon laquelle deux choses sont favorisées à des degrés qui sont à parité est loin d'être clair. Cela représente manifestement une difficulté considérable dans la mesure où, si nous ne pouvons pas faire sens de la notion de parité de degrés de « favorisation » (*favouring*), nous ne pouvons pas non plus faire sens de la notion de parité préférentielle puisque celle-ci est définie en fonction de la première notion.

Une possibilité qui s'offre à nous consiste à modifier une idée proposée par Ruth Chang dans un contexte axiologique (Chang, 2002a, pp. 141-143). *Mutatis mutandis*, nous pourrions affirmer que les degrés auxquels deux choses sont favorisées sont à parité si et seulement si la différence entre de tels degrés est non nulle, c'est-à-dire si elle a une magnitude; et si elle est impartiale, c'est-à-dire si elle ne privilégie aucune de deux choses. Comme Erik Carlson l'a montré, cette idée semble toutefois présenter certains défauts (Carlson, 2010, pp. 119-128; pp. 120-122). Modifiée de telle sorte qu'elle s'applique au cas actuel, la critique de Carlson consiste à montrer qu'il n'est pas possible de comprendre la notion de différence non nulle et impartiale entre des degrés de « favorisation » sans réintroduire l'idée d'une différence partielle, ou sans réduire la notion de parité préférentielle à celle d'absence de préférence.

Une autre possibilité consiste à caractériser l'idée de parité de degrés de « favorisation » en termes de disjonction de deux relations de préférence standards. Par exemple, nous pourrions dire que les degrés auxquels deux choses x et y sont favorisées sont à parité si l'individu, ou bien favorise x à un degré plus élevé qu' y , ou bien favorise y à un degré plus élevé que x . Cette suggestion est toutefois problématique, parce qu'elle viole l'exigence de neutralité de l'analyse en ce qui concerne le débat substantiel sur la parité. En effet, cette suggestion présuppose que la parité est une relation dérivée d'autres relations standards et réductible à celles-ci. Comme nous l'avons vu précédemment, toutefois, l'analyse en termes d'attitudes appropriées de Gustafsson est censée valoir même si la parité s'avère être une relation non dérivée et irréductible. La conception de la parité préférentielle en question doit donc être écartée¹⁰.

En somme, le défi à relever pour Gustafsson est de fournir une conception satisfaisante de la parité préférentielle. À l'évidence, rien de ce qui a été dit jusqu'à maintenant ne peut démontrer de manière concluante que cela ne peut être accompli. Toutefois, le fardeau de la preuve incombe entièrement à Gustafsson.

Il convient de noter que, dans son article, Gustafsson tente de mettre de côté cette difficulté en alléguant que « l'explication des relations de préférence [...] ne relève pas de la portée de l'analyse de la valeur en termes d'attitudes appropriées » (Gustafsson, 2013a, p. 488)¹¹. Or cette réponse s'avère problématique parce que, comme j'ai tenté de le montrer, la validité du premier argument de Gustafsson repose précisément sur la façon dont les préférences sont caractérisées. Si nous avons une raison indépendante de croire au principe de symétrie des valeurs et des préférences, nous pourrions alors accepter une analyse de la parité axiologique fondée sur une notion indéfinie de parité préférentielle. Ce qui est pourtant en jeu reste précisément la question de savoir si nous devons croire ou non au principe de symétrie des valeurs et des préférences. Le fait que la notion de parité préférentielle demeure inexpliquée constitue une pétition de principe contre Rabinowicz.

À ce stade, on pourrait objecter que si nous ne pouvons faire sens de la parité préférentielle, nous ne pouvons non plus faire sens de la parité axiologique. Encore

une fois, cela ne serait vrai que si nous avions des raisons indépendantes de croire au principe de symétrie des valeurs et des préférences. Mais puisque de telles raisons manquent, et puisqu'il existe une autre façon de faire sens de la notion de parité axiologique, nous devrions simplement en conclure que dans de tels cas il existe des relations de valeur dépourvues de relations de préférence correspondantes. En d'autres mots, à moins que Gustafsson puisse fournir une théorie adéquate de la parité préférentielle, nous n'avons aucune raison de penser que les relations de valeur et les relations de préférence sont symétriques.

LE SECOND ARGUMENT DE GUSTAFSSON

Tel que mentionné précédemment, il existe, selon Rabinowicz, des circonstances où il est permmissible pour un individu d'entretenir des relations de préférence différentes à l'endroit d'une paire de choses (bien que non simultanément). Gustafsson admet une telle possibilité. Le problème tient au fait que son cadre théorique ne permet pas de rendre compte des préférences permmissibles multiples, étant donné que pour chaque relation de valeur, il existe une relation de préférence unique que l'on est tenu d'avoir. Si cela est vrai, le cadre théorique de Gustafsson ne parvient pas alors à rendre compte d'une importante caractéristique de la normativité des préférences, laquelle est prise en compte dans le cadre théorique de Rabinowicz.

Gustafsson s'oppose à une telle conclusion. Il affirme que la thèse selon laquelle son cadre théorique ne peut accommoder la notion de relations de préférence permmissibles multiples découle d'une présupposition illégitime, à savoir que dans tous les cas, pour deux relations de préférence distinctes a et b , et pour deux choses x et y , si l'on est tenu d'avoir une relation de préférence a entre x et y , alors il ne nous est pas permis d'avoir une relation de préférence b entre x et y . Gustafsson croit que cette thèse est fautive et qu'elle doit ainsi être rejetée.

Considérons une situation où deux choses x et y sont axiologiquement à parité. Comme on le sait, Gustafsson croit que, dans de tels cas, il est requis d'entretenir x et y en parité préférentielle. Toutefois, selon lui, il ne s'ensuit pas qu'il n'est pas permmissible d'entretenir également une relation de préférence standard entre x et y .

Étant donné que ni « mieux » ni « pire » ne s'appliquent ici, nous ne sommes pas rationnellement tenus de préférer x à y et nous ne sommes pas rationnellement tenus de préférer y à x . Pourtant il pourrait nous être permis rationnellement de préférer x à y et permis rationnellement de préférer y à x , puisque cela est compatible avec le fait d'avoir la relation de préférence rationnellement requise, c'est-à-dire d'entretenir x et y en parité préférentielle en un sens faible (*preferentially weakly on a par*). (Gustafsson, 2013a, p. 490)¹²

Si ce raisonnement est fondé, le cadre théorique de Gustafsson peut effectivement accommoder des relations de préférence permmissibles multiples au même titre que celui de Rabinowicz.

DISCUSSION DU SECOND ARGUMENT DE GUSTAFSSON

Dans ce qui suit, je soutiendrai que l'argument de Gustafsson est incorrect. Convenons d'abord que si deux attitudes sont mutuellement exclusives et que s'il est requis d'avoir l'une d'entre elles, alors il est requis de ne pas avoir l'autre. Cela peut être exprimé de manière plus formelle à l'aide du principe suivant :

(K) À supposer, en tout temps, qu'il y ait deux relations de préférence a et b et deux choses x et y , si l'on est tenu d'entretenir une relation de préférence a entre x et y , et s'il n'est pas possible qu' a et b puissent valoir simultanément (*hold together*), on est alors tenu de ne pas entretenir une relation de préférence b entre x et y .

Manifestement, (K) implique que s'il est requis d'entretenir une relation de préférence a entre x et y , et s'il n'est pas possible qu' a et b puissent valoir simultanément, il ne nous est alors pas permis d'entretenir une relation de préférence b entre x et y . Afin d'avoir des relations de préférence permises multiples, Gustafsson doit alors admettre la possibilité que deux attitudes préférentielles puissent être compatibles. Dans de tels cas, le principe suivant s'applique :

(L) À supposer, en tout temps, qu'il y ait deux relations de préférence a et b et deux choses x et y , si l'on est tenu d'entretenir une relation de préférence a entre x et y et s'il est possible qu' a et b puissent valoir simultanément, il existe alors des cas où il nous est permis d'entretenir une relation de préférence b entre x et y .

La question cruciale consiste à savoir s'il existe des paires de relations de préférence qui ne sont pas mutuellement exclusives. Gustafsson affirme minimalement que la relation de parité préférentielle peut valoir ensemble à n'importe quelle relation de préférence standard. S'il a raison, il existe alors des cas où il est possible d'avoir une relation de préférence non standard requise de même qu'une relation de préférence standard permise.

Gustafsson a-t-il raison? La première chose à constater est que sa solution semble *ad hoc*. Les relations de préférence standards sont (par définition) mutuellement exclusives. Si tel est le cas, pourquoi devrions-nous alors penser que la situation peut être différente lorsqu'il est question de relations de préférence non standards? Il semble n'y avoir aucune raison indépendante de penser que les attitudes préférentielles standards et les attitudes préférentielles non standards ne sont pas mutuellement exclusives. Peut-être Gustafsson affirmerait-il que les attitudes préférentielles standards et les attitudes préférentielles non standards sont compatibles lorsque ces dernières sont dérivées des premières (si, par exemple, la parité préférentielle consiste simplement en une relation de préférence standard indéterminée)¹³. Cette solution est toutefois problématique. Tel que mentionné précédemment, l'analyse en termes d'attitudes appropriées de Gustafsson est censée demeurer neutre en ce qui concerne le débat substantiel sur la parité. Cela signifie qu'en principe nous pourrions accepter l'analyse en termes

d'attitudes appropriées de Gustafsson sans prendre position sur la question de savoir lequel des arguments pour ou contre l'existence de la parité est vrai. Pourtant, la définition de la parité préférentielle examinée, en l'occurrence, présuppose l'inexactitude des arguments en faveur de la parité, c'est-à-dire qu'elle présuppose que la parité est simplement une relation dérivée. Comme telle, l'analyse en termes d'attitudes appropriées de Gustafsson ou bien n'est pas neutre en matière de questions substantielles, ou bien n'est pas à même de rendre compte des préférences permmissibles multiples lorsque la parité est conçue comme une relation authentique, non dérivée.

Or, il est possible que cette critique soit injuste. Après tout, Gustafsson affirme ne pas être concerné par la parité au sens *strict*, mais par la parité au sens *faible*. Autrement dit, Gustafsson affirme seulement qu'il est possible pour un individu d'entretenir deux choses en parité préférentielle *faible* et d'entretenir une préférence stricte à leur égard ou une attitude d'indifférence à leur endroit. Pour évaluer cette thèse, il faut comprendre en quoi consiste une relation de parité faible. Dans son texte, Gustafsson soutient que la parité au sens faible doit être considérée comme une notion préthéorique. Plus précisément, nous avons un cas de parité axiologique faible si aucune des relations de valeur standards ne s'applique et, en même temps, si x est (par exemple) légèrement mieux qu' y , mais pas strictement mieux. Je considère cela, au pire, incohérent ou, au mieux, comme un plaidoyer pour l'introduction d'une nouvelle relation axiologique, notamment une relation d'égalité approximative. Si cela est ce que Gustafsson veut dire, alors, conformément au principe de symétrie entre les valeurs et les préférences, il devrait simplement conclure qu'il y a une autre relation de préférence non standard que l'agent est requis d'avoir, c'est-à-dire l'indifférence approximative. Toutefois, cela n'implique pas qu'il est également permmissible pour l'agent de préférer strictement x à y . S'il n'y a pas de raisons positives de penser autrement, le fait que x est légèrement meilleur qu' y ne rend pas permmissible d'avoir une relation de préférence stricte entre x et y , en plus de la relation de préférence non standard requise (c'est-à-dire, la relation d'indifférence approximative), si aucune des relations de valeur standards ne s'applique entre x et y .

En fait, il y a une raison additionnelle et plus sérieuse de penser que l'argument de Gustafsson ne peut être ultimement concluant. Comme nous avons pu le constater, son argument dépend de la thèse selon laquelle il est possible pour un agent d'avoir simultanément une attitude préférentielle standard et une attitude préférentielle non standard. Supposons, pour les besoins de la discussion, que ce soit effectivement le cas. Il en résulte alors une conséquence importante : la symétrie entre les relations de préférence et les relations de valeur ne tient plus la route. Considérons à présent le cas de la parité (faible). Selon Gustafsson, en ce qui concerne la préférence, on peut à la fois entretenir deux options x et y en parité préférentielle (faible) et, par exemple, être indifférent entre x et y . D'un point de vue axiologique, toutefois, la parité (faible) constitue un cas où aucune des relations de valeur standards ne s'applique. Cela signifie que la parité axiologique (faible) ne peut s'appliquer simultanément avec l'égalité axiologique.

Si cela est vrai, alors les relations de valeur et les relations de préférence se comportent plutôt différemment. Dans certains cas, un individu peut avoir deux relations de préférence distinctes envers une paire de choses, bien que les relations de valeur correspondantes entre ces choses ne puissent valoir ensemble (*hold together*). En effet, si nous définissons une nouvelle relation de préférence c comme la conjonction de la relation de parité préférentielle (faible) et la relation d'indifférence, nous pouvons immédiatement constater que le principe de symétrie entre les valeurs et les préférences ne peut plus être valide, car il n'existe pas de relation de valeur correspondante à c . Par conséquent, Gustafsson fait face à un dilemme : ou bien il abandonne le principe de symétrie entre les valeurs et les préférences, ou bien il admet qu'il ne peut pas rendre compte des cas où différentes relations de préférence sont également permises. D'une manière ou d'une autre, son analyse s'avère problématique¹⁴.

NOTES

- ¹ Traduction de Denis Courville, revue par l'auteur.
- ² Rabinowicz développe, et modifie substantiellement, une ligne de pensée proposée à l'origine par Gert (2004). Puisque la théorie de Gert est sujette à des objections considérables (Chang, 2005, pp. 331-350; pp. 337-344; Rabinowicz, 2008, pp. 30-36), dans ce qui suit je me concentrerai seulement sur l'analyse de la valeur en termes d'attitudes appropriées de Rabinowicz.
- ³ En effet, le principe de symétrie des valeurs et des préférences implique qu'il y a une relation de préférence non standard, la parité préférentielle, qui correspond à la parité axiologique. Cela est en contraste avec l'idée de Rabinowicz selon laquelle il n'existe pas de relations de préférence non standards.
- ⁴ Cela n'implique pas que Gustafsson ne reconnait pas l'existence du niveau du « permisible ». En fait, il est commun de définir la notion de « requis » en termes de la notion de « permisible » (et vice-versa), en disant qu'une chose (ex. une action) est requise si et seulement si son absence n'est pas permisible. Toutefois, la thèse de Gustafsson est qu'il est possible d'analyser toutes les relations de valeur (standards et non standards) sans faire appel à l'idée que certaines relations de préférence sont *tout simplement* permissibles (*merely permissible*).
- ⁵ En guise d'appui supplémentaire à cette ligne de pensée, Gustafsson ajoute que les auteurs qui ont d'abord avancé l'argument de la faible amélioration et l'argument de l'enchaînement (respectivement de Sousa, 1974 et Chang, 2002b) ont soit formulé ces arguments en termes d'attitudes appropriées plutôt que de valeur, soit explicitement maintenu que ces arguments valent autant pour les relations de préférence que pour les relations de valeur.
- ⁶ Voir Gustafsson, 2013a, p. 485.
- ⁷ Pour un argument influent selon lequel les préférences devraient être conçues comme des jugements de valeur, voir Hausman, 2012.
- ⁸ Il vaut la peine de noter que selon certains auteurs (ex. McDowell, 1985; Tappolet, 2000), il est préférable de concevoir l'analyse de la valeur en termes d'attitudes appropriées comme une simple élucidation conceptuelle, plutôt que comme une analyse réductive. Pour des raisons d'espace, toutefois, dans ce qui suit je laisserai de côté cette complication.
- ⁹ J'utilise cette conception des préférences uniquement à titre d'exemple. Mon argument s'applique aussi aux autres conceptions des préférences qui ne réduisent pas les préférences à des jugements de valeur (ex. la conception des préférences en tant que dispositions dyadiques de choix, c'est-à-dire, dispositions de choix vis-à-vis deux options).
- ¹⁰ La proposition discutée dans ce paragraphe est d'ailleurs incompatible avec un autre principe auquel Gustafsson souscrit dans son article, à savoir le principe de la non irrationalité des contraintes axiologiques (*non-irrationality of axiological constrainedness*), selon lequel si aucune relation de valeur ne s'applique entre deux choses, alors il n'est pas rationnellement requis d'avoir l'une ou l'autre des relations de préférence correspondantes. Or, comme nous l'avons vu, les cas de parité axiologique sont précisément des cas où aucune des relations de valeur standards ne s'applique. Le problème est que, si nous définissons la parité préférentielle en termes de disjonction de deux relations de préférence standards, il s'ensuit que dans les cas de parité axiologique, il est requis d'avoir l'une ou l'autre de ces relations de préférence standards. Cela va à l'encontre de ce que le principe de la non irrationalité des contraintes axiologiques affirme.
- ¹¹ Ma traduction.
- ¹² Ma traduction.
- ¹³ Comme Gustafsson le suggère ailleurs (Gustafsson, 2013b, pp. 433-445).
- ¹⁴ Je souhaiterais remercier Krister Bykvist, Erik Carlson, Denis Courville, Jens Johansson, Christopher Kelly, Graham Oddie, Niklas Olsson-Yaouzis, Jean-Charles Pelland, Wlodek

Rabinowicz, Daniel Ramöller, Frans Svensson, Jennifer Szende, Christine Tappolet, et deux évaluateurs anonymes de *Les ateliers de l'éthique* pour leurs précieux commentaires sur des versions précédentes ou des présentations de cet article.

BIBLIOGRAPHIE

- Carlson, Erik, "Parity Demystified," *Theoria*, vol. 76, no. 2, 2010, pp. 119-128.
- Carlson, Erik, "The Small-Improvement Argument Rescued," *The Philosophical Quarterly*, vol. 61, no. 242, 2011, pp. 171-174.
- Chang, Ruth, "Introduction", in R. Chang (dir.), *Incommensurability, Incomparability, and Practical Reason*, Cambridge, Harvard University Press, 1997, p. 1-34.
- Chang, Ruth, *Making Comparisons Count*, New York, NY, Routledge, 2002a.
- Chang, Ruth, "The Possibility of Parity," *Ethics*, vol. 112, no. 4, 2002b, pp. 659-688.
- Chang, Ruth, "Parity, Interval Value, and Choice," *Ethics*, vol. 115, no. 2, 2005, pp. 331-350.
- Danielsson, S. et J. Olson, "Brentano and the Buck-Passers," *Mind*, vol. 116, no. 463, 2007, pp. 11-522.
- de Sousa, Ronald, "The Good and the True," *Mind*, vol. 83, no. 332, 1974, pp. 534-551.
- Espinoza, Nicolas, "The Small Improvement Argument," *Synthese*, vol. 165, no. 1, 2008, pp. 127-139.
- Gert, Joshua, "Value and Parity," *Ethics*, vol. 114, no. 3, 2004, p. 492-510.
- Gustafsson, Johan, "Value-Preference Symmetry and Fitting-attitude Accounts of Value Relations," *The Philosophical Quarterly*, vol. 63, no. 252, 2013a, pp. 476-491.
- Gustafsson, Johan, "Indeterminacy and the Small-Improvement Argument," *Utilitas*, vol. 25, no. 4, 2013b, pp. 433-445.
- Gustafsson, Johan et Nicolas Espinoza, "Conflicting Reasons in the Small-Improvement Argument," *The Philosophical Quarterly*, vol. 60, no. 241, 2010, pp. 754-763.
- Hausman, Daniel, *Preference, Value, Choice, and Welfare*, Cambridge, Cambridge University Press, 2012.
- Hsieh, Nien-Hê, "Equality, Clumpiness and Incomparability," *Utilitas*, vol. 17, no. 2, 2005, pp. 180-204.
- McDowell, John, "Values and Secondary Qualities," in Ted Honderich (dir.), *Morality and Objectivity*, London, Routledge, 1985, p. 110-129. Traduit en français par Ruwen Ogien, « Valeurs et qualités secondes », in R. Ogien, (dir.), *Le Réalisme Moral*, Paris, Presses universitaires de France, 1999, pp. 247-271.
- Peterson, Martin, "Parity, Clumpiness and Rational Choice," *Utilitas*, vol. 19, no. 4, 2007,

pp. 505-513.

Rabinowicz, Wlodek, "Value Relations," *Theoria*, vol. 74, no. 1, 2008, pp. 18-49.

Rabinowicz, Wlodek, "Value compared," *Polish Journal of Philosophy*, vol. 3, no. 1, 2009, pp. 73-96.

Rabinowicz, Wlodek, "Value Relations Revisited," *Economics and Philosophy*, vol. 28, no. 2, 2012, pp. 133-164.

Rabinowicz, Wlodek, et Toni Rønnow-Rasmussen, "The Strike of The Demon: On Fitting Pro-Attitudes and Value," *Ethics*, vol. 114, no. 3, 2004, pp. 391-423.

Scanlon, Thomas, *What We Owe To Each Other*, Cambridge, Harvard University Press, 1998.

Tappolet, Christine, *Émotions et valeurs*, Paris, Presse universitaires de France, 2000.

RICHNESS THEORY: FROM VALUE TO ACTION¹

GREGORY M. MIKKELSON

UNIVERSITÉ MCGILL

ABSTRACT:

Richness theory offers a promising axiology. In this paper, I discuss how to translate it into a deontology. To do so, I recruit the concept of moral distance from a recently developed epistemology, and construe it in terms of causal power. Finally, I apply the resulting decision-theoretic framework to the question of how best to avert ecological disaster over the next 36 years and achieve ecological harmony over the next 986.

RÉSUMÉ :

La théorie de la richesse représente une axiologie prometteuse. Dans le présent article, nous explorons comment la traduire en déontologie. Pour ce faire, nous mobilisons le concept de distance morale tiré d'une épistémologie récemment développée et l'interprétons en termes de pouvoir causal. Pour terminer, nous appliquons le cadre théorique-décisionnel qui en résulte à la question de comment mieux éviter la catastrophe écologique au cours des 36 prochaines années et atteindre une harmonie écologique d'ici 986 ans.

How can a man be satisfied to entertain an opinion merely, and enjoy it? [...] Action from principle – the perception and the performance of right – changes things and relations; it is essentially revolutionary, and does not consist wholly with anything which was. (Thoreau 1849)

DEEP SPACE, DEEP TIME, DEEP VALUE

Ethics, finally, has its Copernicus, its Darwin. What those scientists achieved in 1543 and 1859, Kelly attained in 2003: a comprehensive inference to the best explanation that put to rest one long-standing theory, and replaced it with one clearly more accurate. Though it took decades for many to acknowledge it, heliocentrism replaced geocentrism in 1543. The same goes for evolution *via* natural selection vs. creation *via* intelligent design in 1859. In 2003, richness theory (RT) superseded sentientism as the truest theory of value².

One thing RT explains especially well is the failure – despite repeated attempts by many ethicists – to draw a defensible line between that which has intrinsic value and that which lacks it (see Routley and Routley 1980, Davison 2012, and Mikkelson, forthcoming). RT tells us no such line exists, since everything has intrinsic value. How much intrinsic value anything has depends on how internally varied and internally harmonious it is. RT also resolves the long-standing debate over individualism vs. holism in environmental ethics, in favour of holism (Mikkelson and Chapman, forthcoming). And it thereby squares the idea that an individual human being has greater intrinsic value than just about any other organism, with the merits of a global economy seriously down-sized for the sake of our fellow species (Mikkelson 2011b).

All of the above virtues lie within the domain of value. In this paper, I extend the discussion to the realm of action. I do so through a meta-ethical framework that co-evolved in mutualistic association with RT: robust moral realism (Oddie 2005). I relate one of that framework's most interesting features – its concept of moral distance – to decision-theoretic consequentialism (cf. Jackson 1991). Furthermore, I apply the latter to the question of how best to avert ecological disaster over the next 36 years, and achieve ecological harmony over the next 986.

VALUE, POWER, AND DESIRE

Our whole life is startlingly moral. There is never an instant's truce between virtue and vice. (Thoreau 1854)

Oddie (2005) offers a compelling defense of “robust moral realism”: the idea that the good – i.e., richness – is an objective, causally efficacious, knowable feature of the world³. Much of the book concerns epistemology: how we come to know richness. But its last chapter sows the seed of an action theory, i.e., how we should respond to richness. Here, after sketching the epistemic story – in

particular, the prominent role of desire in that story – I shall critique its action-theoretic implications and argue that decision-theoretic consequentialism addresses the main concern with them.

Like other moral realists (e.g., Boyd 1988), Oddie draws upon an analogy with scientific realism. Science aims to discover various aspects of “reality: the way things really are” (Oddie, 2005, p. 214). “Appearances: the way things seem” provide evidence about reality. Though an individual appearance may be an illusion, other appearances, combined with logical reasoning, expose many such illusions as such. Often enough, this results in “beliefs: the way [one]... believe[s] things to be” that more closely match reality. Likewise, ethics aims to discover “value: how good things really are” (Oddie, 2005, p. 219). And as the moral analogue of appearances in science Oddie identifies desires: “how good things seem”. As with appearances in science, in ethics systematic reflection on desires yields, often enough, “judgements: how good [one]... judge[s] things to be” that more closely match real value.

So far, so good. But one major dis-analogy between science and ethics complicates the picture, and segues the present discussion from value to action. As Oddie points out, in many cases our perceptual experiences (“appearances”) resist any substantial modification in light of our scientific knowledge. For example, coming to know that the sun is in fact much larger than the moon does not change the visual impression, from Earth, that the two are about the same size. But in ethics, we do expect our desires to change as we develop more accurate knowledge of value. This is because desire plays a crucial role not only in coming to know the good but also in responding to it. In order to promote the good, one must first summon the desire to do so.

Oddie seems to acknowledge this obligation to change our desires in light of what we know about value:

“[O]ur value judgments [...] can and often do track value [...]” “Those judgments in turn can play an important [...] role in shaping [...] desires [...] An accurate theory of the good may correct for, or eliminate, a wayward desire.” (Oddie, 2005, p. 218)

But he insists that considerable gaps between value and desire should nevertheless often remain. Rather than tracking just the real value of things, our desires should also reflect how close we are, in moral space, to those things. Oddie provides a few hints about what moral distance might involve, yet they fall short of dispelling the worry that the value-desire gaps he endorses are moral failings rather than virtues.

Facts about relative power can dispel that worry. Each of us stands in a better position to promote certain goods than other goods. To follow up on one of Oddie’s own examples, you plausibly stand in a better position to promote the recovery of your own spouse from “a very serious, life-threatening illness” than to help out a total stranger “in a village in some far-flung country” whom “[y]ou know nothing about [...] and have no connection with [...] other than that they share your spouse’s illness” (Oddie, 2005, p. 221). More generally, given a choice

between two goods with equal objective value, but greater power to promote the first one than the second, a person should prefer the first good to the second one – i.e., desire the former more strongly than the latter. I submit that this – and only this – construal of moral distance can justify Oddie’s claim that even after coming to know that “two states of affairs [...] are of equal value, the one that is closer to you should [nevertheless still] exert greater motivational force on you” (Oddie, 2005, p. 223).

The idea that relative power, as well as value, should affect our desires and thus our actions fits well with decision-theoretic consequentialism (DTC; see Jackson 1991). According to DTC, an agent should choose the course of action with the highest expected value. The expected value of an act is a function of not just the values, but also the probabilities, of that act’s potential outcomes. I propose to interpret the kind of power discussed in the previous paragraph as *the probability of success in achieving an intended outcome*. If this outcome has greater value than the probability-weighted mean value of the alternative and unintended outcomes of the action, then the greater the chance of succeeding, the greater the expected value of trying. (See the appendix for a proof.) This brings the idea of moral proximity espoused by Oddie into rough alignment with the concept of expected value propounded by Jackson. And it supports the above analysis of moral distance in terms of causal power.

HUMANITY’S ECOLOGICAL FOOTPRINT

All human agents, together, face a real choice now between futures with vastly different richness, and thus value. Overwhelming scientific evidence shows that the current path of ever-increasing depletion and pollution of nature will lead inexorably to planetary social and ecological collapse. This is not news. Already in 1972 the authors of *The Limits to Growth* demonstrated the disturbing implications of sustained exponential growth, such as the sixteenfold expansion undergone by the global economy in the twentieth century (Meadows *et al.* 1972). But so far the ruling class, who require such absolute growth to augment their relative power, have stymied environmentalists’ efforts to put us on a better course^{4 5}.

The expected number of premature human deaths from global warming alone – just one of the “interacting travesties” resulting from this business-as-usual (BU) approach (Mikkelson 2011a) – runs to the billions (Broome 2012). Orders of magnitude more untimely non-human deaths – and perhaps millions of premature extinctions – are also expected (Thomas *et al.* 2004, Mikkelson and Chapman, forthcoming). Species extinction ties in closely with richness theory, since it reduces both the variety and the harmony of the living world (Mikkelson 2011b). But so does the death of individuals, since organisms, and especially humans, are also prime examples of richness (Kelly 2003).

The Global Footprint Network (GFN) call their preferred alternative to BU the rapid-reduction scenario (RR; GFN 2014). In it, humanity immediately begins “climbing down” from its current demand for 1.5 Earths’ worth of renewable resource supply and pollution assimilation to reach one Earths’ worth by 2050. Even this measure would provide no guarantee against global ecological havoc.

For one thing, it would be crazy and evil for humanity to use even “just” one Earth’s worth of “biocapacity” for very long: crazy, because it would leave zero margin for error; and evil, because it would leave little or nothing to the ten million wild species who contribute the lion’s share of the global ecosystem’s overall richness (Mikkelsen 2011b)⁶. For these reasons, it will be imperative to continue relaxing the human footprint after 2050 – for example, by slowly reducing it to around 10% of biocapacity by the year 3000 (Nash 2011).

Nonetheless, the GFN’s RR scenario would surely be a good start. It would make human die-offs and non-human extinctions much less likely to occur, and much less severe if they do occur, than the GFN’s BU projection, which entails a global ecological footprint of between two and three Earths’ worth by 2050.

I hope the above comments suffice to establish that the value (i.e., the richness) of RR vastly exceeds the richness of BU. But what about the probabilities? BU is, by definition, the path we’re already on. Thus, if we aimed to stay this course, we would arguably have a greater chance of success than if we dug in our heels and fought instead for RR. To use Oddie’s (2005) language, RR lies further away from us in moral space than does BU.

However, causal analysis of the global ecological footprint (EF) – perhaps the most comprehensive measure of environmental damage currently available – shows that RR is not as improbable, and hence distant, as it might seem. The relative EF per unit of economic output has already been declining – by some 64% from 1961 to 2008 (data from the GFN, United Nations, and World Bank)⁷. Projecting forward this rate of improvement in socio-technical efficiency means that humanity could reach a one-Earth EF by 2050 – if growth in human population size and per-capita production and consumption slowed down enough. Jointly sufficient conditions would be for population growth to ease up from its recent rate of 1.7%/year to around 0.6%/year and for growth in per-capita production/consumption to slacken from 1.9%/year to 0.6%/year. In other words, global economic growth need not stop altogether – though it must slow down tremendously.

What – or, more to the point, who – stands in the way of realizing the RR scenario? Many academics suppose that the general public are too blinkered to appreciate such esoterica as the intrinsic value of richness in nature. But surveys refute this all-too-common form of condescension. For example, Leiserowitz *et al.* (2005) report that overwhelming majorities of Americans agree with such statements as “Nature has value within itself regardless of any value humans place on it” and “Humans have moral duties and obligations to plants and trees”. On the other hand, the wealthiest minority do have great trouble understanding ecological values and duties. Studies show very consistently that the proportion of a person’s own income that he or she is willing to pay for environmental protection declines precipitously with increasing income (Jacobsen and Hanley 2009)⁸.

I conclude that although it will certainly be harder to achieve RR than BU, the far greater value of RR greatly outweighs this difference in difficulty. The expect-

ed value of aiming for RR greatly exceeds the expected value of settling for BU, despite the lower probability of success in achieving RR, because the richness of RR so vastly exceeds that of BU. We should thus strongly prefer RR to BU – i.e., desire the former much more strongly than the latter, despite the fact that RR lies further away from us in moral space than does BU – because, again, the value of RR so vastly surpasses that of BU. To act on this preference, as we should, calls for direct efforts to reduce growth in human population size and per-capita production and consumption⁹. Given the intransigence, alluded to above, of those who currently hold the most wealth and power, the transition to a just and sustainable future also requires a massive shift of “power to the people”.

APPENDIX

The expected value (E) of an action is the sum (Σ), over all n possible outcomes of that action, of the probability (p) of the outcome, given the action, multiplied by the outcome's value (G):

$$E = \Sigma_{i=1 \text{ to } n} p_i G_i$$

If we identify one of those outcomes as the intended one (I), and the others as unintended, then the formula becomes

$$E = p_I G_I + \Sigma_{u=1 \text{ to } n-1} p_u G_u$$

The probability-weighted mean value of the alternative, unintended outcomes (V_F , i.e., the expected value of failing to achieve the intended outcome) is

$$V_F = (\Sigma_{u=1 \text{ to } n-1} p_u G_u) / (1 - p_I)$$

Thus,

$$E = p_I G_I + (1 - p_I) V_F$$

Consider two possible actions A and B, the intended outcomes of which have equal value ($G_{IA} = G_{IB}$), and the unintended outcomes of which have equal probability-weighted mean value ($V_{FA} = V_{FB}$), but of which A has the greater probability of success in achieving its intended outcome ($p_{IA} > p_{IB}$).

Ex hypothesi (see discussion in main text above),

$$G_{IA} > V_{FA}$$

I need to prove that, given the foregoing assumptions,

$$E_A > E_B$$

Or equivalently, that

$$E_A - E_B > 0$$

Proof:

By definition,

$$E_A = p_{IA}G_{IA} + (1 - p_{IA})V_{FA}$$

By substitution,

$$E_B = p_{IB}G_{IA} + (1 - p_{IB})V_{FA}$$

Therefore, after re-arrangement of terms,

$$E_A - E_B = (G_{IA} - V_{FA})(p_{IA} - p_{IB})$$

Since, by assumption, $G_{IA} > V_{FA}$ and $p_{IA} > p_{IB}$, both factors on the right-hand side of this equation are positive. Their product is therefore also positive and thus,

$$E_A - E_B > 0$$

NOTES

- ¹ I thank Peter Caines, Jeff Mikkelson, Jean-Philippe Royer, Christine Tappolet, Adam Westra, Tiger Zheng, two anonymous reviewers for feedback on this paper. For helpful questions and comments on the talk the paper is based on, I thank fellow attendees of the workshop on “Attitudes, Values, and Environment” held at the University of Montreal on February 21, 2014; and especially Graham Oddie for pointing out again to me the affinity between my construal of moral distance and decision-theoretic consequentialism.
- ² Of course, just as Copernican heliocentrism and Darwinian evolution had antecedents reaching far back into history, so too does RT, e.g., with Leibniz (1686) and Hutcheson (1725).
- ³ Though one need not accept richness theory in order to accept robust moral realism, pp. 130, 132, and 140 make clear that Oddie himself does subscribe to the former view.
- ⁴ Zinn (2003) and Schweickart (2011) ascertain the ruling class in the US as roughly the wealthiest 1% of the population – the same 1% whose domination the 2011-2012 Occupy movement explicitly challenged.
- ⁵ Powell (1971) laid out the blueprint for this anti-environmental counter-movement, shortly before his appointment to the US Supreme Court, where he personally helped advance one important means for thwarting socio-ecological progress, namely new rights for corporations as “persons” under the Constitution. See Schweickart (2009) for an explanation of why the 1% have so far locked us into our ecocidal growth-at-all-costs trajectory.
- ⁶ Many wildlife populations are already collapsing due to humanity’s excessive demands on nature. For example, tiger populations fell by an average of 70% from 1980 to 2010 (WWF 2012).
- ⁷ By the “relative” EF I mean the raw global EF, divided by global biocapacity. See GFN (2014) for details.
- ⁸ This fact is a special case of a much broader and very well established pattern: As a proportion of income, the wealthy are stingier donors toward not only environmental quality, but also charitable causes in general; and this, both in terms of their stated willingness to pay as well as their actual donations (Piff *et al.* 2012).
- ⁹ The best ways to achieve these goals are to greatly improve, respectively, world-wide access to birth control (Engelman 2011), and workers’ freedom to scale back their work hours – and thereby income and consumption – without losing their jobs or benefits (Knight *et al.* 2013).

REFERENCES

- Boyd, Richard N., *How to be a moral realist*, in Geoffrey Sayre-McCord (ed.), *Essays on Moral Realism*, Ithaca, NY, Cornell University Press, 1988, pp. 181-228.
- Broome, John, *Climate Matters*, New York, W. W. Norton, 2012.
- Copernicus, Nicolaus, *De Revolutionibus Orbium Coelestium*, Nuremberg, Germany, Johannes Petreius, 1543.
- Darwin, Charles, *On the Origin of Species by Means of Natural Selection*, London, John Murray 1859.
- Davison, Scott A., *On the Intrinsic Value of Everything*, New York, NY, Continuum, 2012.
- Engelman, Robert, "An end to population growth: Why family planning is key to a sustainable future," *Solutions*, vol. 2, no. 3, 2011.
- Global Footprint Network (GFN), www.footprintnetwork.org/en/index.php/GFN/page/world_footprint, 2014.
- Hutcheson, Francis, *An Inquiry into the Original of our Ideas of Beauty and Virtue*, oll.libertyfund.org/titles/hutcheson-an-inquiry-into-the-original-of-our-ideas-of-beauty-and-virtue-1726-2004, [1725].
- Jackson, Frank, "Decision-theoretic consequentialism and the nearest and dearest objection," *Ethics*, vol. 101, no. 3, 1991, pp. 461-482.
- Jacobsen, Jette Bredahl and Nick Hanley, "Are there income effects on global willingness to pay for biodiversity conservation?," *Environmental and Resource Economics*, vol. 43, no. 2, 2009, pp. 137-160.
- Kelly, Crhistopher, *A Theory of the Good*, Ph.D. Thesis, University of Colorado, Boulder, CO, 2003.
- Knight, Kyle W., Eugene A. Rosa, and Juliet B. Schor, "Could working less reduce pressures on the environment? A cross-national panel analysis of OECD countries, 1970-2007," *Global Environmental Change*, vol. 23, no. 4, 2013, pp. 691-700.
- Leibniz, Gottfried W, *Discourse on Metaphysics*, in R. S. Woolhouse (ed.), *Philosophical Texts, G. W. Leibniz*, Oxford, Oxford University, [1686] 1998, pp. 54-89.
- Leiserowitz, Anthony A., Robert W. Kates, and Thomas M. Parris, "Do global attitudes and behaviors support sustainable development?," *Environment*, vol. 47, no. 9, 2005, pp. 22-38.
- Meadows, Donella H. *et al.*, *The Limits to Growth*, New York, NY, Universe Books, 1972.
- Mikkelson, Gregory M., "Diversity and the good," in B. Brown, K. de Laplante, K. Peacock, D. M. Gabbay, P. Thagard, and J. Woods (eds.), *Handbook of the Philosophy of Science. Volume 11: Ecology*, Toronto, ON, Elsevier, 2011a, pp. 399-415.
- Mikkelson, Gregory M., "Weighing species," *Environmental Ethics*, vol. 33, no. 2, 2011b, pp. 185-196.
- Mikkelson, Gregory M., "On the intrinsic value of everything," *Environmental Ethics*, forthcoming.
- Mikkelson, Gregory M. and C. A. Chapman, "Individualistic environmental ethics: A *reductio ad extinctum*?," *Environmental Ethics*, forthcoming.
- Nash, Roderick F, "Island civilisation: A vision for human occupancy of Earth in the Fourth Millennium," in P. Burdon (ed.), *Exploring Wild Law*, Kent Town, Wakefield Press, 2011, pp. 339-348.

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

Piff, Paul K. *et al.*, "Higher social class predicts increased unethical behaviour," *Proceedings of the National Academy of Sciences*, vol. 109, no. 11, pp. 4086-4091.

Powell, Lewis F., Jr., "Attack on American Free Enterprise System," URL: <http://law.wlu.edu/deptimages/Powell%20Archives/PowellMemorandumPrinted.pdf>.

Routley, Richard and Val Routley, "Human chauvinism and environmental ethics," in D. S. Mannison, M. A. McRobbie, and R. Routley (eds.), *Environmental Philosophy (Monograph series / Department of Philosophy, Research School of Social Sciences, Australian National University)*, Atascadero, CA, Ridgeview Publishing Co., 1982, pp. 96-189.

Schweickart, David, "Is sustainable capitalism an oxymoron?," *Perspectives on Global Development and Technology*, vol. 8, no. 2-3, 2009, pp. 559-580.

Schweickart, David, *After Capitalism*, Lanham, MD, Rowman and Littlefield, 2011.

Thomas, Chris D. *et al.*, "Extinction risk from climate change," *Nature*, vol. 427, 2004, pp. 145-148.

Thoreau, Henry D., "Resistance to civil government: A lecture delivered in 1847," in E. Peabody (ed.), *Æsthetic Papers*, New York, NY, AMS Press Inc., [1849] 1967, pp. 189-213.

Thoreau, Henry D., *Walden, or Life in the Woods*, Boston, MA, Ticknor and Fields, 1854.

Worldwide Fund for Nature (WWF), *Living Planet Report*, Gland, Switzerland, WWF International 2012.

Zinn, Howard, *A People's History of the United States: 1492 – Present*, New York, NY, Harper-Collins, 2003.

VALUE MONISM, RICHNESS, AND ENVIRONMENTAL ETHICS

CHRIS KELLY
INDEPENDENT SCHOLAR

ABSTRACT:

The intuitions at the core of environmental ethics and of other neglected value realms put pressure on traditional anthropocentric ethics based on monistic value theories. Such pressure is so severe that it has led many to give up on the idea of monistic value theories altogether. I argue that value monism is still preferable to value pluralism and that, indeed, these new challenges are opportunities to vastly improve impoverished traditional theories. I suggest an alternative monistic theory, Richness Theory, and show how it provides an opportunity to capture the needs of both environmental ethics and of our traditional ethics.

RÉSUMÉ :

Les intuitions au cœur de l'éthique de l'environnement et celles d'autres domaines négligés de valeurs exercent une pression sur l'éthique anthropocentrique traditionnelle basée sur les théories monistes de la valeur. La force importante de cette pression fait que plusieurs ont abandonné ces théories. Nous soutenons que le monisme des valeurs demeure préférable au pluralisme des valeurs et que, en fait, ces nouveaux défis offrent la possibilité de grandement enrichir les théories traditionnelles appauvries. Nous proposons une différente théorie moniste, la théorie de la richesse, et montrons en quoi elle permet de capter les besoins à la fois de l'éthique de l'environnement et de l'éthique traditionnelle.

INTRODUCTION

Environmental ethics is a relatively new field of study, but it is built on ancient intuitions about the value of natural objects, the value of life, the value of natural beauty, and the value of the remarkable interdependence of nature. During the industrialization of the Western world, these old intuitions have clearly been neglected both in the culture at large as well as in the predominant academic ethics. In particular, there is little place for those intuitions in the traditional person-entered ethics that have dominated Western philosophy over the past several centuries. These ethics have sought to maximize human well-being in the form of pleasure or happiness or preference satisfaction. Alternatively, they have sought to protect the rights of human beings: the right to life, property, well-being, and so on. The development of these theories has run parallel to a general growth in the well-being of the Western citizen. These theories certainly arose, at least in part, as a response to the socio-political challenges of the past few centuries. Many of these challenges have lessened in the Western world, and new challenges have arisen—chief among them is a rapidly worsening environmental crisis. Unfortunately, the ethics that arose to improve human lives are ill-suited to capture the intuitions mentioned above. Ecosystems have no feelings. The oceans do not (obviously) have rights. Although some animals probably do suffer and feel versions of pleasure and happiness, can we naturally extend some notion of rights to them? And once we do, how do we still account for the strong intuition that a human life counts for more than that of a mouse, for instance? The general response is that natural objects are only instrumentally valuable, but this is deeply unsatisfying to many of us. Surely, there is something good in itself in a rain forest: its beauty, its richness, the depth of its species, etc. The rainforest's existence makes for a better world, all other things being equal, and this is so independently of whether or not the forest does any good to humans to at all. By contrast, the dominant ethical theories are generally very stingy with their ascription of intrinsic value. Indeed, for reasons of parsimony, the underlying value theories are usually monistic. Things like knowledge, beauty, autonomy, and love also seem like things that are valuable in themselves, but this value is hard to capture using theories that focus exclusively on human well-being.

I propose that most modern-day theories fail to take seriously the multiple presentations of value in our world. I will argue, nonetheless, that the draw of monism is justified for well-established reasons of theoretical simplicity. Yet the benefits of simplicity do not obviate the need to explain the evidence. I present a quick formulation of a monistic theory that does take the value evidence seriously. I do not give a full defense of this theory, but I do hope to show its plausibility for the present investigation. Finally, I show that this theory can naturally account for the traditionally difficult value areas at the center of environmental ethics.

MONISM VS. PLURALISM

Value pluralism is the thesis that there are many basic (i.e., non-reducible), intrinsically valuable properties in the world. Value monism is the thesis that there is only one intrinsically valuable property. Pluralist theories of value have strong advantages over existing monistic theories because they more easily capture the value evidence. On the face of it, we value many things: we love people and paintings; we like foods and generosity; we care about flowers and physics; we value autonomy and pleasure. We treat many of these things as ends in themselves. If all of these things are intrinsically valuable, then the monist must show that they all share some single property that accounts for their degree of value. This is a serious obligation—one that, most would admit, has yet to be met by any monistic theory. And so there are great evidentiary reasons to be a pluralist about value.

There are two types of pluralist to consider; call them the radical pluralist and the moderate pluralist. The radical pluralist claims that there are many (at least two, though usually many more) different values and that these values are incommensurable. (Two measures are incommensurable if it doesn't make sense to compare them, for example, my body temperature and my IQ.) The moderate pluralist claims that all values are commensurable, but not reducible.

Radical Pluralism

There are a number of well-regarded arguments undermining radical pluralism. Indeed, I have argued elsewhere that radical pluralism is impossible¹.

I will give a brief overview of that argument here.

A widely-accepted basic feature of value is that it should give rise to motivations². As Mackie put it: “An objective good would be sought by anyone acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness built into it³.”

A proper appreciation of value, then, must motivate us. Furthermore, the strength of these motivations should be commensurable with the motivating value. A more valuable thing should (normatively) motivate us more than a less valuable thing. Graham Oddie couples these two assumptions in what he calls the Merit Connection⁴:

(MC) State of affairs S merits a desire with a strength proportional to the value of state of affairs S.

These two widely-held assumptions lead to the conclusion that all values are mutually commensurable. If the strengths of motivations are commensurable (and no one has suggested the contrary), and if all values are commensurable to some desire strength, then all values are commensurable to each other.

This result should be quite welcome. It is a well-known feature of radical pluralism that it makes rational decision theory impossible. For this reason alone, many philosophers have found value incommensurability undesirable. Of course, this is an extremely brief dismissal of a commonly defended position. I touch on it here merely to orient the reader with respect to the structure of the larger argument within which this paper is situated.

Moderate Pluralism

The moderate pluralist asserts that there is an objective, non-reducible, commensurable plurality of values. This is an extremely plausible starting point, as there does seem to be a plurality of not obviously reducible things that we value. Indeed, it is this piece of value evidence that demands explanation from the environmental ethicist's point of view. It is also very close to our folk theory of value: we all pursue multiple values.

Because these plural values are commensurable, there must exist a set of proportionality measures between them that preserve the transitivity of value. Suppose x , y , and z are the fundamental values. To make this a complete value theory, we need to supplement it with something like the following: (a) $x=2y$, (b) $y=2z$, and (c) $x=4z$. So our theory of value is that x , y , and z are valuable and (a), (b), and (c) hold between them.

What explains this complicated set of value facts? What, for example, is the explanation for (a) $x=2y$? It is not an analytic truth. Even supposing goodness be defined in terms of some natural property x , the value of x will not be part of the analysis of x . Consider: Good =def, x =def (analysis of x). If x is supposed to serve as a naturalistic reductive definition of x , then no value terms can occur in (analysis of x); that is, it cannot be part of the definition of x that x is twice as valuable as y . But this should be unsurprising: does anyone find it plausible that value proportions should be part of the final analyses of life, love, happiness, pleasure, freedom, knowledge, rightness, or any of the other prospective values?

But the only other plausible explanation for (a) is not available to the pluralist, namely that x has twice as much of some good property P as y . This option is monistic: it reduces the value of x , y , and z to the value of P . In the end, the pluralist has stipulated away the only possible explanation for (a). And so the pluralist is forced to accept a number of, in principle, inexplicable synthetic truths—not just the basic facts that x , y , and z are intrinsic non-reducible goods, but also the facts that the measures between them are what they are. Clearly, this becomes more and more unpalatable the greater the number of values we postulate; so a theory like Michael Stocker's, which posits nearly as many types of value as there are experiencers, becomes, on the face of it, little more than a description⁵. Most of the facts in such a 'theory' of the good remain unexplained.

Let's briefly consider an analogous theoretical situation from the field of physics. In nineteenth-century physics, the concept of force was front and center (indeed, it may still be). A force is a "vector quantity"; it has both direction and magnitude. At that time, scientists were exploring the 'ecology' of forces by measuring them against each other using balances, pulleys, and the like. Newton had given us the gravitation force (and his three laws of motion). We had the electrostatic force. We had magnetism. But there also seemed to be chemical forces, and frictional forces, tension, and others. Indeed, before Newton, there were innumerable force classes: forces from 'my hand' and from 'your hand', from 'rivers smashing against dams', from 'tides', from 'fire', and so on. In the twentieth century, two other fundamental forces were discovered in the quantum world: the weak force and the strong force.

This situation is analogous to the situation in which we find ourselves with respect to value. Scientists realized that all of these 'types' of forces could be balanced against each other, measured—they were 'commensurable.' They had a common unit of measure (called a 'Newton'). On the other hand, the sources of 'force' seemed innumerable and, in many cases, not at all similar. Force appeared, then, to be plural and commensurable. Of course, that *could have* been how the world was ordered; there may have been no further underlying order or unification. But the same considerations that I have outlined above drove scientists to seek an explanation for this ecology of forces. And the fact that these forces could all be measured and balanced against each other was a strong encouragement to seek out a monistic theory of force. That project has been wildly productive (if not yet fully successful). By the end of the nineteenth century, electricity, magnetism and (the vast varieties of) light were all unified in the electromagnetic force. By the middle of the twentieth century, all forces had been reduced to four: gravity, and the electromagnetic, weak, and strong forces. At present, there are only two forces resisting unification: gravity and the forces unified by Grand Unified Theories (electroweak and strong). Moreover, much of the progress in physics during the past 150 years, both theoretically and practically, has flowed from the drive to unify this apparent plurality.

In any theory of the world, there will, undoubtedly, remain something at least partially unexplained. And if no 'ought' can be derived from an 'is,' any value theory will have one or more partially unexplained basic facts. But the unexplained thoughts entailed by moderate pluralism do not have the feel of basic facts; they seem top beg for explanation. What makes the good attach to x exactly twice as much as to y ? In general, if two things are commensurable, the explanation for this fact is that they share some property. The reason why the temperature of my hand and the temperature of a Miami day can be compared is that they both share a property: heat, or kinetic energy. The reason why I can compare my weight with that of Jupiter is that they both share a property: mass. Presumably, the reason why we can compare my intelligence with yours is that we both share some property: intelligence. If the values x , y , and z are all commensurable, then we immediately have a good reason to assume that there is some property P shared by x , y , and z that makes each of them valuable to a

certain degree. P is an explanation for the commensurability of x , y , and z . But the pluralist is committed to claiming that there is no explanation for their commensurability.

On the face of it, this is a strange claim to make. What reasons in general can philosophers have for giving up on the search for explanations besides claiming to have arrived at a good one? A monist has a principled reason to stop seeking an explanation. If property P is the only good, it must be that P's being the good is, at least partially, basic and unexplained; either because P is good by definition, or by synthetic necessity. The pluralist, by contrast, is under the obligation to provide some similar justification for giving up his exploration of value. The only one forthcoming is that she has tried hard to find a unified value theory and has failed.

But the plausibility of this explanation clearly depends on how hard she has tried. I, of course, think we haven't tried hard enough: the number of monistic theories of the good that have been vigorously explored can probably be counted on one hand, and have focused, for the most part, on properties closely tied to human psychology, such as pleasure, desire-satisfaction, and reason. Shouldn't we look a bit further afield? After all, if value is part of the structure of the objective universe, why need it be captured by some obvious, easy to grasp, human-oriented concept? The basic concepts of physics certainly aren't so easily grasped, and we have yet to find a unified theory of our physical universe. But are we justified in giving up?

Pluralism is what I would call a philosophically unstable theory; it will always leave us with the suspicion that there is something deeper to be discovered. If ever we did arrive at a good pluralist theory, we would (and should) still be severely tempted to continue looking for the unity underneath those plural values. And so without some argument to show that monism is impossible, we have good reason to continue the search. Moreover, if we find a monistic theory that can capture as much evidence as a pluralistic one, the monistic theory is the better one.

There is a second reason why an environmental ethicist should be wary of the pull of pluralism: assuming a pluralist stance makes persuasion particularly difficult. The practical urgency behind environmental ethics is that things of fundamental value are being neglected and destroyed by cultures bound to traditional anthropocentric ethics. Convincing people that there are plural intrinsic values will do nothing to convince them that ecosystems are intrinsically valuable unless they already have that intuition. There is no further argument one can make to convince them of the value of ecosystems, non-human animals, or natural objects. Pluralists often have the feeling that they are creating more space for respecting value. But the reverse is more likely to be true. If I do not see the value of ecosystems in the first place, there is nothing to which a pluralist can appeal in order to convince me of their value. This is one reason why so many environmental ethicists return to the monistic anthropocentric the-

ories when making persuasive arguments: only in this way can they appeal to common values to convince others of the value of important natural objects.

Ultimately, if we can formulate a monistic theory that captures the values of environmental ethics *as well as* the more obvious human-centered values, we will be in a much better position to convince those who doubt the value of ecosystems, animals, etc. This consideration, by the way, is not merely practical. If we are defending values that cannot be shared with, or shown to, large numbers of other valuers, we have a good reason to doubt that we are defending objective values. Indeed, the ability to persuade an open-minded interlocutor through appeals to reason and evidence is a type of functional evidence that we are on strong theoretical grounds. This surrender of certain forms of pluralist on the battlefields of persuasion is often the first and last step back to the subjectivity or relativity of value because it suggests there is no independent fact of the matter to function as the ground for persuasion.

Accordingly, the claim is that, the ideal value theory has the following structure: *there exists a property P whose distribution in the world determines the value structure of the world.* This is either because the property P is identical with the property of being good, or because the property P is the only thing that has the property of being good.

All things being equal, a monistic theory that can do a reasonably good job of capturing the plurality of value evidence will be preferable to any pluralistic theory. As it stands now, however, pluralistic theories are more viable because they are more accurate. We manifestly value love, life, art, memory, and many other things; and while pluralism can easily capture this diversity, monism has had great difficulty doing so. Because of this difficulty, traditional monistic theories have resorted to *explaining away* the evidence rather than explaining it.

Indeed, pluralism enjoys an advantage over monism to the extent that it takes our evidence seriously and attempts to explain it, whereas monistic theories, such as early utilitarianism, almost invariably attempt to explain away much of our value evidence. Suppose we are eudemonistic utilitarians. Our general method of explaining away is to claim that any intuition we might have that some V other than happiness is valuable is in fact due to V's being normally instrumental to creating happiness, i.e., that there is some strong, but contingent connection between V and happiness. Our intuitions that life, integrity, honesty, not-killing, beauty, knowledge etc. are valuable are all explained away in this manner. They only appear to be valuable because they are often useful for creating happiness. Not surprisingly, many of the strong objections to utilitarianism charge that it fails to take seriously our intuitions that things other than happiness are valuable.

So what do we do? We want to be monists, but we must capture the evidence of value pluralism. We are in the same situation as any good scientist. By inference to the best explanation, we must search for a unifying explanation of all the things we value. If love, beauty, freedom, knowledge, ecosystems, people, and happiness all appear to be intrinsically valuable, then they must all share some value-giving property.

VALUE AS RICHNESS

What could this possibly be? In a book-length study, I defended the thesis that a property I call ‘richness’ is an excellent candidate to be the property that unifies the apparent plurality of value⁶. While I do not have space to fully make that argument here, I want to show how assuming monism, coupled with freeing ourselves from the shackles of anthropocentric ethics, can move us quickly into new theoretical territory.

If we take seriously that all value comes from the same source, then we are freed up to take seriously value realms outside of traditional ethics. One plausible place to start would be aesthetics. Aesthetics is, as Nozick put it, “the area that speaks most frequently and articulately about value⁷.” There are innumerable accounts of aesthetic value, of course. Fortunately, only a few purport to be monistic and objective. Moreover, few plausibly point at something that might be able to apply outside of aesthetics, that is, that might undergird a broader value theory.

As far back as Plato, philosophers and artists have argued that beauty is equivalent to ‘organic unity,’ as it has most commonly been called. The concept of organic unity originated when philosophers noted the similarity between good art and living creatures. Plato said of rhetoric that “all the parts should be in keeping with each other and with the whole like a living being, with a body of its own, so as not to be headless or footless, but so as to have middle and members, composed in keeping with each other and with the whole⁸”.

The organic metaphor has been in use ever since. An artist tries to create things as natural and surprising as the multifarious living world. There are many features of living beings to which the organic metaphor might be pointing. First, consider the property of having tightly interdependent parts: “Men call beautiful the things in which the parts fully answer to each other [...]” said Dante. The functioning of most of the parts of a living body depends on the further functioning of most of the other parts. A human being needs its liver to survive; the liver needs the heart to function; the heart needs the lungs to function; and the lungs need the liver. This interdependence also characterizes the best art. In living beings, the parts must answer to each other. What the heart asks for, the liver gives; what the liver gives, the heart takes; what the liver asks for, the heart gives; what the heart gives, the liver takes. Likewise, in the best art no part is placed without respect for the effect it will have on the other parts or without an understanding of how it itself will change in the context of the other parts.

This should catch our interest. Already, we see two disparate value realms being connected, aesthetics and living creatures. Indeed, this particular connection is especially interesting from the standpoint of environmental ethics: interdependence is not only a feature of living organisms, but also of ecosystems—one of the more difficult value entities to defend with individualistic ethics.

A number of attempts have been made to capture this notion of interdependence philosophically. Aristotle said of tragedy that it is a whole, “the structural union of the parts being such that, if any one of them is displaced or removed, the whole will be disjointed and disturbed. For a thing whose presence or absence makes no visible difference, is not an organic part of the whole¹⁰.” As John Hospers put it: “In the unified object, everything that is necessary is there, and nothing that is not necessary is there [...] in a work of art, if a certain yellow patch were not in a painting, its entire character would be altered, and so would a play if a particular scene were not in it, in just the place where it is¹¹.”

Dewitt Parker claims that ‘the ancient law of organic unity’ is “the fact that each element in a work of art is necessary to its value, that it contains no elements that are not thus necessary, and that all that are needful are there¹².” And as Harold Osborne put it, “[t]he theory of organic unity claims that any subtraction or addition would diminish the value of the work (of art) as a whole, changing also the character of all the parts [...]”¹³.

This ‘ancient law’ is a common method of evaluating works of art. Few, if any, artistic works meet this ideal of unity, of course. Even the greatest works have filler words and passages that presumably could have been replaced with other words and passages without causing harm to the work. What Dante, Aristotle and the others are pointing to is the ideal artistic work. All art, they are saying, strives to be an organic unity.

So interdependence is one attempt to capture the notion of organic unity. The other common attempt claims that an organic unity is a unified variety.

Francis Hutcheson defined beauty as a “compound ratio of uniformity and variety: so that where the uniformity of bodies is equal, the beauty is as the variety; and where the variety is equal, the beauty is as the uniformity.”

Harold Osborne in his *Theory of Beauty* says, “All works of art, even the simplest, are fairly complex for immediate perception and it has been a commonplace of artistic criticism that a good work of art must have a high degree of unity... [...] The two notions, however vague, suggest that works of art of whatever kind may have in common the formal property of being *complex unities*.”

In the *Ars Poetica*, Horace says that a poem should be a unity of different parts. Coleridge repeatedly cited “unity in multiplicity” as his formula for aesthetic success. John Hospers has stated that the organically unified object “should contain within itself a large number of diverse elements, each of which in some way contributes to the total integration of the unified whole, so there is no confusion despite the disparate elements.” Santayana identified “unity in multiplicity” as the key measure of the value in aesthetic form.

Robert Nozick also argued that most value is reducible to the value of organic unity. Obviously, we need a value property that can come in degrees, he said: “Holding fixed the unifiedness of the material, the degree of organic unity varies

directly with the degree of diversity of that material being unified. Holding fixed the degree of diversity of the material, the degree of organic unity varies directly with the degree of unifiedness [...]”¹⁴⁹.

So let’s see if we can put these two concepts of organic unity together. I call the resulting property ‘richness.’ According to Nozick and the unity-and-variety camp, richness is a measure of the unity and variety in a thing. We could capture it thus:

$$R = UV$$

This formula captures important boundary conditions. A state of affairs with no variety, with nothing going on inside, will be minimally rich. Consider a musical piece consisting of a single note or a painting that is nothing but a blank canvas. If we ignore any meta-richness (for instance, the richness hidden in the commentary about music that something like ‘4:33’ by John Cage exhibits), these are minimally rich objects because they have minimal variety. Moreover, these are not impressive works of art. Likewise, a completely unordered, chaotic state of affairs will have minimal richness. An artwork with no structure or cohesion, a work that is purely random, will have minimal richness. In aesthetics (and elsewhere), we are looking for an ordered variety, something complicated brought together harmoniously.

But what do we mean by unity and variety?

In analyzing unity, it pays, I suggest, to consider closely related, if not synonymous, terms, such as order, simplicity, uniformity, harmony, and coherence. What do these concepts have in common? All of them imply a *limitation* on the number of properties in a state of affairs. This is straightforwardly true of simplicity and uniformity. But the same is true for unity and order. A whole becomes unified in virtue of its parts’ sharing certain properties. One way of putting this is that the number of *determinates* exhibited for a particular *determinable* is minimized. A determinate is a specific way of being a determinable; e.g., scarlet and carmine are both determinates of the determinable red in that they are ways of being red. A collection of Americans is a somewhat ordered state because, regardless of all the differences among the individuals in the collection, they all share a property, namely, that of being citizens of the United States. The determinable of *citizenship* is fixed, limited to one determinate, and this is what the order of that collection consists in. The more determinables we fix or limit in our collection, the more ordered it becomes. A determinable need not be limited to only one determinate for some state to be unified. A house that is red and blue is unified by that colour scheme: the determinable colour is limited to only two of its many determinates. A maximally unified state is one in which all the parts of the collection share all their properties. Presumably, this state consists of something like one indivisible concrete particular (or a collection of concrete particulars indiscernible from each other). A minimally unified state consists, presumably, in a large collection of objects that share as few properties as possible. Let us suppose, then, that the unity axis of richness is a measure of the *limitations* on the number of determinates instantiated for some

determinable.

An increase in variety, on the other hand, is a *reduction of limitations* on determinables. The more varied a state is, intuitively, *the greater the number* of determinates exhibited for each determinable. The most varied state with respect to the determinable ‘hair colour’ is one that presents all possible determinates, i.e., all possible hair colours. A collection of Americans that exhibited all colours of hair would be more varied than one that presented only two colours of hair. And a collection of persons that exhibited all possible citizenships and all possible hair colours would be a more varied state than one that exhibited only Americans of many different hair colours. If this is right, variety and unity vary inversely. Variety is lack of unity, and unity is lack of variety¹⁵.

This causes serious problems, however. UV, so construed, won’t be a good measure of anything, because UV will always equal 1 (or some other constant). Variety will never increase without a corresponding decrease in unity and vice versa. Perhaps, then, (in defense of Hutcheson’s and Nozick’s take) variety and unity within richness should be maximized over different determinables. For instance, a collection of Americans all with the same hair colour is unified under two determinables and varied to some extent under all others. On this account, if variety were increased in those other determinables, we would increase the richness. This move could preserve a form of the original definition. For some whole **W** that exhibits two sets of determinables **U** and **V**, richness is that property that increases in whole **W** when the unity with respect to determinables **U** increases while the variety with respect to determinables **V** remains constant, *or* that property that increases in whole **W** when the variety with respect to some determinables **V** increases while the unity with respect to some determinables **U** remains constant. So:

$$R_w = U_u V_v$$

Good. The question at this point, though, is which determinables to put in set **U** and which in set **V**.

There are a number of ways to go from here, but I would like to continue with our general approach. If monism were true, then a good monistic theory would have to explain all objective value. Say, for the sake of argument, then, that our brief overview of the organic unity position in aesthetics hasn’t sharpened our concept of richness sufficiently. We could turn elsewhere in the value realm and gather more evidence. By way of seeing how this approach would work, let us investigate another set of widely held values neglected by traditional anthropocentric ethics.

Theoreticians share a set of strong and consistent value intuitions. Indeed, we have already appealed to those intuitions in this paper. We strongly prefer simple theories to less simple theories. And we strongly prefer simple theories that capture lots of evidence to simple theories that capture less evidence. This is not

just an intuition of epistemic accuracy, since it is notoriously difficult to cash out this intuition as an epistemic justification. We just find ‘elegant,’ ‘powerful’ theories more beautiful and more desirable. I suspect that even if it were shown that parsimonious theories were no more likely to be true than less parsimonious theories, we theoreticians would still be strongly tempted to prefer the former. The value preference for simplicity is widespread and consistent. It has as much claim to being an objective value as most any. In addition, there is surprisingly little controversy over what it means in practice, even though the notion of simplicity is not easy to define (simply). We can generally agree on which of two theories is the more parsimonious.

Despite its ubiquity, this strong intuition of value has rarely been taken seriously *qua* value. If we are monists about value, though, we *must* take it seriously. We must attempt to explain this value and unify it with our other conceptions of value. It is obvious that most anthropocentric value theories have no chance to do this. How could the value of human well-being and theoretical parsimony ever be captured by the value of a single property?

One philosopher who did take the theoretical values seriously was Leibniz:

One can say that he who acts perfectly is like an excellent geometer who knows how to find the best constructions for a problem; or like a good architect who utilizes his site and the funds destined for the building in the most advantageous manner, leaving nothing which offends or which falls short of the beauty of which it is susceptible; like a good paterfamilias who puts his capital to use in such a way that nothing is left waste or barren; like a skilful engineer who makes his effect by choosing the least difficult way; like a talented author who encloses a maximum of realities in the least possible volume. Now the most perfect of all beings and those which occupy the least possible volume, that is to say which hinder each other least, are spirits, and their perfections are the virtues [...]

God has chosen the [rule] which is the most perfect, that is to say the one which is at the same time the simplest in hypotheses and the richest in phenomena, as a geometrical line might be, of which the construction was easy and the properties and effects very admirable and of great extent¹⁶.

Here, Leibniz raises to the highest level the value of simple rules giving rise to powerful consequences. This is a perfect universe because it is built out of the most theoretically powerful rules.

This gives us a possible route out of the problem of the inverse nature of unity and variety. Instead of just accounting for the overall variety and/or unity in a whole, we need to account for particular types of unity and particular types of variety. Elsewhere I have argued that we can turn Leibniz’s intuition into a definition of the following form¹⁷:

$$R_w = U_{\text{explanation}} V_{\text{explanandum}}$$

The richness of **W** is the variety in the object multiplied by the unity of the best explanation for the variety of the object. This will not work, though, unless we only allow for certain types of explanations. This is because the best explanation for many, if not most, states of affairs would lie in the causal history leading to that state of affairs, whereas, that the actual causal history of **W** shouldn't be necessary for determining the richness of **W** (the Mona Lisa, for example, would be a rich object even if it popped into existence randomly). I suggest that because I am appealing to explanations to help *unify* explananda—because what is wanted is something that captures the *patterns* in the explananda—we need *nomie* explanations, i.e., explanations that look like laws.

Start with a simple example. The following famous series can be seen as unified by the fact that it suggests a *nomie* explanation:

(F) 1 1 2 3 5 8 13 21 34 55 89 144 233 377 610 987 1597 ...

There are many explanations for how these numbers might appear on the page before you. Explanations about how the printer copied them from the galleys, or how I learned them in discrete mathematics in junior high school, or some story about an Italian man named Fibonacci, or an account of the history of the Golden Ratio. But none of these explanations serve the present need; none of them are *nomie* in nature. Here, though, is a possible *explanatory law*: if y is in the $(n-1)$ ordinal position of (F) and x is in the $(n-2)$ ordinal position, then the number in the n^{th} position of (F) will equal $x+y$. This is a law-like explanation of all elements of the series other than the first two. Moreover, *it is precisely the reason why we see this series of numbers as unified*.

Now for a significantly more complicated example. James Joyce's novel *Portrait of the Artist as a Young Man* was, for its time, radically structured. The novel progresses in voice and tone throughout the body of the work, gradually passing from a child-like voice to the voice of an intelligent, rebellious young man. Many of the traditional unities of the novel are broken: tone, diction, even subject matter to a large extent. I want to suggest that the reader accepts this variety because she largely understands the reason for it, that is, she senses a law-like explanation for this variety. The reader grows up with the protagonist, and the language does as well. The law is therefore something like: if the protagonist is of age x , then he will write like someone of age x . What we ask for in an artwork (and in richness) is not just variety, but a non-arbitrary variety. We need a reason for the variety, an explanation for the way the artwork is. And what kind of explanation are we looking for? It is closely connected to the artist's intentions, of course, but, again, a good novel should not depend for its enjoyment on our knowing what the actual genesis of the words was. Rather, we want to know if the novel, by itself, is intelligible. I suggest that there is a method for explaining each of Joyce's choices, a method that does not require our know-

ing, say, that he happened to be drinking coffee while he wrote chapter four. We need only discern the ‘laws’ of the book, the rules governing the choices. These rules are grammatical, semantic, representational, dramatic, among others no doubt. They are for the most part not deterministic, and the *actual* rules that Joyce used are, of course, somewhat inaccessible to us. But, nonetheless, the unity we require of an artwork is just the sense that there *are* rules governing the choices—that there are structural reasons for why the novel is the way it is, for why the particular words are where they are. In the end, we want the sense that the work has some unitary explanation.

We look for law-like explanations of the variety in a rich object just as theoretical physicists look for law-like explanations of the variety of phenomena that makes up our world (which is also, we suspect, a rich object). To be unified, recall, is to follow *explanatory* rules. It may be that *Portrait* obeys the following rule: (R1) There shall be no less than ten instantiations of the letter ‘e’ on any one page. But this rule fails to be explanatory; it is *ad hoc*. Nothing about the fact of ten e’s on each page makes us think that such a rule played any role in the genesis of the book. It is, therefore, not a properly unifying property of the book. On the other hand, if every word starts with a ‘p,’ then we are tempted to think that that was no accident, that it could only be explained by someone or something’s following a rule entailing such a pattern.

This distinction between an explanatory rule and a non-explanatory rule gives us one method for understanding Aristotle’s claim (and those of Leibniz and Coleridge) that in rich objects “the whole is prior to its parts”; that is, it allows us to capture the notion of organic unity as ‘interdependence.’ It is plausible to think that rich wholes are wholes in virtue of the rules governing them. Without those rules, the parts would *not* be parts of any whole because they would not be unified by anything. “Only when the whole has been dissolved will they [the parts] attain actuality,” says Aristotle¹⁸. Explanations are prior, in some difficult-to-define sense, to what they explain. Non-explanatory rules, by contrast, are not prior to what they attempt to explain (hence the sense that they are *ad hoc*, that they are added on afterward). And so an explanatory unification is prior to the variety it explains.

This is a very brief attempt to get a little sharpness around the concept of richness. I have gone further with this analysis elsewhere, but this will suffice for the present essay. It should be noted that even the present account is a sharper definition than anyone in the aesthetics literature has attempted—sharper, also, than the definition Nozick used to defend a similar monistic thesis. Here I want to show that taking monism seriously gives us new avenues of investigation and that these avenues show real promise. By taking monism seriously, we are forced to be both more creative and more critical. A true monist about value should have had a very difficult time, excepting a Benthamite ethics, for example.

We’ve already seen that the theory that richness is the good is a promising aesthetic theory and almost an exact account of the values core to good theorizing. We’ve also seen hints that it might be able to capture the value of living things.

I have argued elsewhere that it is also a promising account for the value of autonomy, knowledge, experience, originality, love, virtue, happiness (of a certain kind), persons, preference satisfaction, and a number of other things¹⁹. The argument is that all of these things are rich, and that our natural intuitions of which of them are better than others correspond to our intuitions about which of them are richer than others.

For this paper, though, let us see how richness provides a promising account of environmental value.

ENVIRONMENTAL ETHICS

Let us look to three areas of trouble for the environmental ethicist. Anthropocentric theories have trouble capturing the appropriate value of non-human animals, the value of ecosystems, and the value of (untouched) wilderness.

Let's go through them one by one. I will, with the environmental ethicist, assume that these things have value. The obligation for the richness theorist will then be to show that each of these things is rich and that it is their very richness that gives them their value.

Non-human Animals

The prodigious *diversity* of macroscopic structures of living beings rests in fact on a profound and no less remarkable *unity* of microscopic makeup²⁰. [italics mine]

Life is a paradoxical phenomenon. It is enormously *varied* – with creatures ranging from one thousandth of a millimeter to dozens of meters in length, having life-spans from hours to thousands of years, and with some species that spread over the whole globe, while others stay in their tiny ecological niche. And still, life is extremely *uniform*. All cells are built according to the same principles from the same molecular building blocks, no matter whether they are free-living microbes or a tiny part of a huge organism, and regardless of their position in the big family tree of life, [...] ²¹. [italics mine]

Living organisms are the prototypical examples of richness. The variety of functioning structures in a full-grown animal is a true wonder. There are over two hundred different cell types in the human body. Each living cell has a variety of different structures within it. A 'typical' plant cell, for instance, contains, at least, the following sub-structures: Golgi apparatus, endoplasmic reticulum, vacuole, vesicle, ribosome, cell wall, chloroplast, cell membrane, nucleus, mitochondrion, cytosol, middle lamella. And complex clumps of cells form bigger structures, such as tissues and organ systems. More fundamentally, there are twenty amino acids that regulate the functioning of each cell. These amino acids link together in uncountable combinations to compose the larger proteins that form the basic structures of the cell and, in turn, of the organism as a whole. The

way in which the amino acids link together determines how they function; the order of the amino acids and the shape of the final protein distinguish the proteins from each other. It is estimated that there are 5,000 different proteins expressed in each cell.

One could go on and on: the structure of DNA, the variety of life-states an organism undergoes over time, the behavioural responses to the multitude of stimuli an organism can receive.

And yet all of this is clearly unified to a great degree. The unity is made apparent by each organism's consistent structure over time, its set of regular behaviours, and the fact that each organism contains similar underlying biological 'instructions'—the DNA—that appear in and govern each and every living cell. We can identify a single organism through time just by following it; it occupies a unique shape in space-time. We can also identify a single individual by its peculiarities of behaviour. And each living organism has a unique set of DNA that governs its everyday functioning and overall development; the fact that all the events that make up living a history can be explained, at least partially, by a single set of instructions suggests remarkable unity.

The structures and behaviours of every living creature are largely what they are because they have been useful in the continuance of that organism's genes. The incredible variety of functions in each living creature all work together so that nearly all aspects of the creature can be understood as leading toward the end goal of healthy reproduction. This suggests that some rule like S: "P is in organism O if P contributes to reproduction and overall survival of O in environment E."

This is not as simple a rule as it looks, of course. S masks the hidden complexity in the rules that govern E, the laws of physics and chemistry. Nonetheless, given a set of rules governing a specific environment, one can expect a *nomic* explanation for why most of the parts of a living organism function the way they do. These rules are determined by the 'purpose' of survival and the regularities of the environment from which the creature evolved and in which it lives. An animal will breathe on a regular basis as long as the air is not noxious. This breathing requires the regular functioning of multiple organs and muscle groups, which in turn requires the proper functioning of countless cells. All of this is intelligible as part of a method for maintaining the energy required to do other things necessary for survival, such as gathering food, avoiding predators, and so on. The entire structure, more or less, of a living organism is intelligible in this way. It is an incredibly complex and incredibly harmonious thing—and, therefore, incredibly rich.

This, then, gives us a strong reason to think that richness could account for the value we generally give living beings. Does this help with the animal rights question?

Most of us have strong intuitions that non-human animals deserve some moral respect. Indeed, the intuitions are strong enough that anthropocentric theories

have been somewhat successful in stretching their notions of value to include non-human animals. Rights-oriented theories have first reduced rights to having interests and then argued that animals have interests. Somewhat more plausibly, in my opinion, maximizers of preference or eudemonia have argued that animals have preferences and/or suffer pain and must therefore be accounted for in the maximizing calculations.

The theory that richness is the good certainly can explain the intuition that non-human animals are intrinsically valuable and, therefore, intrinsically deserving of moral respect. But our intuitions of the value of animals are more specific than the simple proposition “animals have intrinsic value.” It is not just that we think animals matter, it is that we think certain animals matter more than others and that humans matter the most. Only the most extreme of us would rate the rights, suffering, and life of a mouse as highly as those of a human. But the individualist theories of animal rights can’t account for this intuition. In fact, they generally attempt to *explain away* such feelings as speciesism. But accounting for a natural speciesism does not eliminate this intuition. No serious person is considering bringing mice into the benefits of the welfare state, for example. But, if a mouse has an interest in being alive, then it has exactly the same (inviolable) right to life as a human. If a mouse can suffer pain, its pain counts just as much as the pain of a human being.

So it is plausible that richness can account for the value of living beings. More importantly, maybe, it can capture the further and more specific intuition we have that certain living beings are more valuable than others. If so, it will have a leg up on traditional accounts of animal rights.

A natural way to capture our intuitive ranking of the value of various species is according to the species’ apparent complexity. We value a mineral less than a plant, less than a worm, less than mouse, less than a bear, less than a human. Gregory Mikkelson has argued that one good richness proxy in this debate is the number of cell types a species has, cell types being able to stand in for the variety of functions and abilities an animal might have. “This richness proxy affirms that humans are special. Most if not all non-human organisms have fewer cell types than humans,” says Mikkelson²². This does not mean that animals are not worthy of moral respect, of course, nor that humans will win every conflict between animal interests and human interests. Merely, our intuitive ranking of animal value matches our intuitive ranking of the richness of these animals. This strong correlation is powerful evidence that what we value in living beings *qua* living beings is their richness²³. And, importantly, it gives us an account of animal value that does not fall prey to the problems that arise for traditional extensions of anthropocentric ethics.

Ecosystems

One of the deepest divides within the environmental ethics community is between holists and individualists—those who think that things like ecosystems have intrinsic value versus those who think that any value ecosystems have derives from the individuals affected by that ecosystem. This divide clearly maps

onto those who have tried to extend anthropocentric ethics to account for environmental value versus those who feel that such ethics will always be inadequate. For this reason, there have been few, if any, environmental ethics that can establish the intrinsic value of ecosystems and still support the traditional intuitions of human-centered ethics.

The theory that richness underlies all value can bridge this divide. Clearly, an ecosystem possesses richness. Ecosystems are massively complex, somewhat unified systems. In many ways, they mirror the structure of living organisms. Intuitively they are not as tightly unified as individual organisms, but they are unified to a degree. It is meaningful to talk about the harmony of an ecosystem. It is meaningful to talk about the interdependence of an ecosystem, how each part plays its 'role' in an overarching pattern. Ecosystems have some variety that individuals don't, the obvious being the variety of different organisms, each playing a role. Conversely, ecosystems lack some of the richness that individual creatures have, such as conscious awareness of the world and self-awareness, to name a couple of important features.

Again, then, we see that conceding value to ecosystems need not trump all human or individual interests. Likewise, conceding most of the tenets of traditional ethics need not require denying the important intrinsic value of ecosystems.

On a related note, it is clear that we also value the diversity of life in and of itself. This is easily explained by the value of richness: a functioning (unified) ecosystem with greater diversity is clearly richer than one functioning equally well, but with less diversity. The loss of a species is a significant decrease in the variety of the biosphere. This is a natural explanation of the value of biodiversity in our ecosystems.

Wilderness

The value of wilderness can also be captured by richness theory. In general, human intervention in an otherwise wild area disturbs the unity of the area. A soda can in the deep rainforest of Brazil is jarring: it doesn't belong there, just as a Metallica guitar solo doesn't belong in the *Ode to Joy*. Wilderness ecosystems have their own unified functioning that is almost invariably harmed by human intervention.

Again, none of these considerations are overriding in themselves. A less diverse but more unified ecosystem may be better than one with greater diversity. The value of timber may override the value lost in clear-cutting and replanting (although this becomes less and less true the more ancient the forests that are destroyed). And there may come a day when we learn to alter ecosystems for their own benefit, a day when we can build homes and lives so well integrated with the wild that they actually increase the beauty and value of the area.

CONCLUSION

The demands of pluralism, both politically and philosophically, put pressure on us to account for the full range of human value. Though this can at first seem theoretically daunting, it is also an opportunity. Indeed, it forces us into the methodological stance with which we should have begun. True value theorists need to give an account of all value. Many apparent problems in ethics and axiology may evaporate when we take account of the full range of evidence available to us. I hope that, at the least, I have shown the promise of such an approach.

It should be also obvious that we are not bound to traditional anthropocentric theories. Indeed, the limitations of these theories are so severe (explaining away or neglecting most of the value evidence) that we should abandon them immediately. This does not mean that we need to abandon the underlying intuitions that made them so useful; rather, we need to incorporate those intuitions into a broader theory. At least one plausible alternative is richness theory. It has been defended in various forms for millennia, and it promises to help us unite the various value realms.

NOTES

- ¹ Kelly, Christopher, "The Impossibility of Incommensurable Values," *Philosophical Studies*, vol. 137, no. 3, 2008, pp. 369-382.
- ² Osborne, Harold, *Theory of beauty; an Introduction to Aesthetics*, London, Routledge & K. Paul, 1952.
- ³ Mackie, John L., *Ethics: Inventing Right and Wrong*, New York, Penguin Books, 1977.
- ⁴ Oddie, Graham, *Value, Reality, and Desire*, New York, Oxford University Press, 2005.
- ⁵ Stocker, Michael, *Plural and Conflicting Values*, Oxford, Oxford University Press, 1990, pp. 267-268.
- ⁶ Kelly, Christopher, *A Theory of the Good*, Ph. D. Thesis, Boulder, CO, University of Colorado, 2003.
- ⁷ Nozick, Robert, *The Examined Life: Philosophical Meditations*, New York, Simon and Schuster, 1989, p. 415.
- ⁸ Plato, *Phaedrus*, edited by A. Nehamas, et al., Indianapolis, Hackett, 1995, 264c.
- ⁹ Dante, A., *Il convivio*, edited by G. Busnelli et al., Firenze, Le Monnier, 1964, Iv3-15.
- ¹⁰ Aristotle, *Poetics*, edited by D. W. Lucas, Oxford, Clarendon Press, 1968, VIII.51A 32-35.
- ¹¹ Hospers, John, "The Organic Conception in Aesthetics," in Paul Edwards (ed.), *Encyclopedia of Philosophy*, New York, Macmillan, 1967, p. 3.
- ¹² Aristotle, *Poetics*, op. cit., VIII.51A 32-35.
- ¹³ Lord, Catherine, "Organic Unity Reconsidered," *Journal of Aesthetics and Art Criticism*, vol. 22, no. 3, 1964, pp. 263-268.
- ¹⁴ Nozick, Robert, *Philosophical Explanations*, Cambridge, MA, Harvard University Press, 1981, p. 416.
- ¹⁵ Catherine Lord makes this point. She says, "[m]aximum unity tends to curtail variety. On the other hand, the greater the variety, the less unity there is likely to be. They vary inversely with one another." Lord, "Organic Unity Reconsidered", op. cit., p. 268.
- ¹⁶ Leibniz, Gottfried W., *Discourse on Metaphysics*, Manchester, Manchester University Press, 1953, p. 9.
- ¹⁷ Kelly, *A Theory of the Good*, op. cit..
- ¹⁸ Aristotle, *Metaphysics*, edited by R. Hope, Ann Arbor, University of Michigan Press, 1966, V.11. 1019a 9-10.
- ¹⁹ Kelly, *A Theory of the Good*, op. cit..
- ²⁰ Monod, Jacques, *Change and Necessity: On the Natural Philosophy of Modern Biology*, New York, NY, Alfred A. Knopf, 1971, p. 142
- ²¹ Gross, Michael, *Life On The Edge: Amazing Creatures Thriving In Extreme Environments*, New York, NY, Plenum Trade, 1998, p. 1.
- ²² Mikkelsen, Gregory, "Weight Species," *Environmental Ethics*, Volume 33, no. 2, 2011, pp. 185-196.
- ²³ "Thus the ranking of organisms in accordance with degree of organic unity matches our value ranking of them, with people above other animals above plants above rocks." Nozick, *Philosophical Explanations*, op. cit., p. 243.

INDIVIDUALIST BIOCENTRISM VS. HOLISM REVISITED*

KATIE MCSHANE

COLORADO STATE UNIVERSITY

ABSTRACT:

While holist views such as ecocentrism have considerable intuitive appeal, arguing for the moral considerability of ecological wholes such as ecosystems has turned out to be a very difficult task. In the environmental ethics literature, individualist biocentrists have persuasively argued that individual organisms—but not ecological wholes—are properly regarded as having a good of their own. In this paper, I revisit those arguments and contend that they are fatally flawed. The paper proceeds in five parts. First, I consider some problems brought about by climate change for environmental conservation strategies and argue that these problems give us good pragmatic reasons to want a better account of the welfare of ecological wholes. Second, I describe the theoretical assumptions from normative ethics that form the background of the arguments against holism. Third, I review the arguments given by individualist biocentrists in favour of individualism over holism. Fourth, I review recent work in the philosophy of biology on the units of selection problem, work in medicine on the human biome, and work in evolutionary biology on epigenetics and endogenous viral elements. I show how these developments undermine both the individualist arguments described above as well as the distinction between individuals and wholes as it has been understood by individualists. Finally, I consider five possible theoretical responses to these problems.

RÉSUMÉ :

Quoique les perspectives holistes telles que l'écocentrisme exercent un attrait intuitif considérable, affirmer la considérabilité morale des tous écologiques comme des écosystèmes s'est avéré une tâche très difficile. Dans la littérature en éthique de l'environnement, certains biocentristes individualistes ont argumenté de manière persuasive qu'un organisme individuel, mais pas un tout écologique, peut correctement être considéré comme possédant son propre bien. Dans le présent article, nous réexaminons ces arguments et soutenons qu'ils sont voués à l'échec. Ce travail est divisé en cinq parties. Premièrement, nous nous penchons sur certains des problèmes que pose le changement climatique pour les stratégies de conservation de l'environnement et affirmons que ces problèmes fournissent de bonnes raisons pragmatiques pour lesquelles chercher une meilleure compréhension du bien-être des tous écologiques. Deuxièmement, nous décrivons les a priori théoriques de l'éthique normative qui sous-tendent les arguments contre l'holisme. Troisièmement, nous réexaminons les arguments de biocentristes individualistes appuyant l'individualisme et rejetant l'holisme. Quatrièmement, nous explorons de récents travaux en philosophie de la biologie sur le problème de l'unité de sélection du biome humain en médecine et des éléments épigénétiques et viraux endogènes en biologie évolutionniste. Nous montrons en quoi ces conceptions minent à la fois les arguments individualistes susmentionnés et la distinction entre l'individu et le tout tel que compris par les individualistes. Finalement, nous considérons cinq réponses théoriques possibles à ces problèmes.

TWO PROBLEMS CAUSED BY CLIMATE CHANGE

Consider two much-discussed problems that climate change has posed for environmental conservation¹. The first problem occurs within ecological restoration. Typically, the aim of restoration has been to undo damage or destruction caused to an ecosystem—to restore, as far as possible, that which was damaged or destroyed to the way it was previously². The benchmark used is typically a historical one³. For example, where I live, at the edge of Colorado's eastern plains, restoration projects often aim to return farmland or rangeland to shortgrass prairie, which is what was there before the farmers, ranchers, and housing developments arrived.

Climate change has made—and will increasingly make—problems for this strategy⁴. First, as the climate changes, what existed in a place before might not be able to survive there anymore. For this reason, restoration done with some kind of historical benchmark might well become much more difficult, perhaps even impossible. Second, ecological restoration was initially taken to be a laudable environmental practice on the assumption that what did well in a certain place in the past is likely to do well there in the future. That is to say, the goals of historical fidelity and ecological welfare coincided⁵. But because of climate change, we can no longer assume this coincidence as we look to the future.

The second problem occurs with species preservation⁶. The assumption used to be that species preservation was best accomplished through the preservation of native habitat—in situ preservation⁷. But as the climate changes, many species will not be able to survive in situ. Those that are not able to move to new, more suitable, locations face extinction. For this reason, some conservation biologists have begun to advocate assisted colonization (sometimes called 'assisted migration' or 'assisted dispersal') as a strategy for species preservation⁸. If we want to preserve a species, and if its members cannot survive where they currently are, then the best solution may be to move them to a place where they will do better. Of course in doing that, we would be significantly changing the ecology of the new location, and it is important to consider whether that would be a change for the better or for the worse. Although previously a concern to preserve species and a concern to avoid ecological harm went hand in hand, now in some cases they seem to be coming apart, and we may need to decide which should take priority in cases where they conflict.

In both cases, aims that used to coincide (historical fidelity and ecological welfare; species preservation and ecological welfare) are beginning to diverge, and conservationists face some difficult questions about what to do. There is a growing literature on this topic in conservation biology⁹, but my concern here is a philosophical problem underlying many of these discussions, namely the question of what we might plausibly mean by 'ecological welfare.' When we ask what kind of ecosystem will do best in the new climate in some location, when we ask whether the introduction of a new species will harm or benefit an existing ecosystem, what exactly are we asking? Is this a question of what is good for an

ecosystem, or is it really a question of what is good for us (or good for sentient beings, or good for individual organisms) with regard to an ecosystem?

The above examples show why this question has practical importance today more than ever. While before we might have been able to use ‘restoration to historical conditions’ or ‘preservation of an endangered species’ as proxies for ‘promotion of ecosystemic welfare,’ we can no longer do so with confidence. This means that we will either have to learn how to talk about ecosystemic welfare directly, or else give up on the idea that there is anything in the world answerable to the idea of the flourishing of ecological communities as such.

Interestingly, while environmental policy seems increasingly to need a conception of ecological welfare, environmental ethicists these days tend to regard the very idea of the good of an ecological whole as a kind of conceptual mistake¹⁰. In this paper, I will revisit the argument that led many to this conclusion in the first place and consider whether it really does rule out the literal attribution of welfare to ecological wholes. In the second section, I review the theoretical assumptions from normative ethics that form the background of these discussions within environmental ethics. In the third section, I describe the central argument that has been given in favour of individualist versions of biocentrism (hereinafter ‘individualism’) and against holist versions of biocentrism (hereinafter ‘holism’). In the fourth section, I argue that recent work in the philosophy of biology, work in medicine on the human biome, and work in evolutionary biology on epigenetics and endogenous viral elements undermine this argument and threaten the distinction between individuals and wholes as it has been understood by individualists. In the final section, I consider five possible theoretical responses to these problems. Ultimately, I argue that the main argument for the superiority of individualism to holism fails. While this does not show holism to be justified, it does remove one of the main obstacles to philosophically justifying the welfare of ecological wholes.

THEORETICAL BACKGROUND

In environmental ethics, the view that nothing can technically be ‘good for’ an ecological whole comes out of a debate within biocentrism between individualism and holism. Biocentrism is the view that all and only living things should be deemed morally considerable. *Individualists* claim that all and only individual organisms should be deemed morally considerable, while *holists* claim that ecological wholes (biomes, species, ecosystems) should also be deemed morally considerable. For a thing to be ‘morally considerable,’ as I am using the phrase here, is for its interests to matter morally in their own right, such that its interests ought to weigh in moral agents’ decisions about what to do¹¹. For the purposes of the current analysis, I will not consider the question of whether one should be a biocentrist at all—though this is a question well worth asking—and for the sake of simplicity I will only discuss one particular ecological whole: ecosystems.

If, as many have assumed, a central part of moral decision-making involves considering how various interests would be affected by different decisions, then an important question for a moral agent to answer is “Which interests do I need to consider?” For the most part, environmental ethicists have taken a broad view about which interests matter: if a thing has interests, then its interests matter¹². And so the important question within environmental ethics has become, “Which things have interests?”¹³ That is, which things have a welfare, or a well-being? Which things are capable of being benefitted or harmed? It is in trying to answer this question that individualists and holists have parted ways. Individualists have argued that only individual organisms have interests—wholes, such as species and ecosystems, do not have interests. They claim that individuals that make up wholes have interests, but the wholes themselves do not.

The main challenge that all biocentrists—holists or individualists—face in extending the notion of interests to cover entities that have no subjective experiences of the world is finding a nonarbitrary basis for the assignment of interests. With a typical human being, we can say that it matters to her how her life goes even if it does not matter to anyone else. She has a unique point of view on the world, and from this point of view, things such as adequate nourishment and the alleviation of pain are properly regarded as good. It is this point of view that serves as the basis for our attribution of interests. We can say that from her point of view the alleviation of her pain is good, even if from everyone else’s point of view the alleviation of her pain is bad. It is not too difficult to extend this notion of interests from human beings to other conscious life forms—dogs, for example. A dog has its own point of view on the world from which the alleviation of its pain is a good thing, regardless of what the rest of us think or what would be good for us¹⁴. This subjective point of view on the world is what grounds our claims that the dog has a good of its own, that what is good for the dog is not just a function of what is good for the rest of us. Proponents of this argument contend, by way of contrast, that we cannot say this about mere ‘things’: what is good for cars is just a matter of what is good for people; cars do not have a good of their own. But dogs do, and people do.

It is, I think, worth reflecting on why this point has been so important in Western ethical thought. In both ethics and political philosophy, taking an instrumentalist attitude toward other people—treating them as though the only interests relevant to how they should be treated are the interests of the State, or the interests of the ruling class, or the interests of those who write moral philosophy—has been considered anathema, an attitude that underlies the worst kinds of immorality: genocide, slavery, eugenics, etc¹⁵. Western ethics has long considered it important to respect individuals in a way that refuses to assimilate their interests to the interests of the wholes of which they are a part. Many writers (e.g., Bentham) have argued that States or other collectives should not even be thought of as having interests, that only individuals have interests¹⁶. The justification for claims such as these is that individuals have a subjective point of view on the world—my interests matter to me, even if they do not matter to the State. A morally acceptable order is one that respects this fact about individuals.

Individualist biocentrists have argued that we can say the same of any individual organism—that even though it does not have subjective states, it still has a good of its own that is not reducible to the interests of others. Let us consider how the argument for that claim proceeds.

THE ARGUMENT FOR INDIVIDUALISM

Biocentrists typically argue that once we understand the function of subjective states in their biological context, we will see that subjectivity is just one of the many tools that organisms use to process information about their environment and to adjust their behaviour accordingly¹⁷. That is to say, subjective or conscious states are just one of the ways that organisms have evolved to pursue their basic biological goals: self-defence, metabolism, reproduction, etc. In humans, for example, pleasure and pain (among other things) represent to us which things in our environment are good (i.e., to be sought) and which things are bad (i.e., to be avoided); pleasure and pain thereby serve as important sources of information for us about our environment. Pleasure and pain states are not simply informative in this regard, they are also motivational—they motivate our seeking and avoidance behaviour. Amoeba also have ways of identifying things in their environment that are to be sought or avoided and initiating the appropriate seeking or avoidance behaviour; they just do it via different mechanisms¹⁸. Our ways of figuring out what is good for us and pursuing it partly rely on mechanisms that utilize subjectivity: internal representations to ourselves of certain aspects of the external world. Other organisms' ways of figuring out what is good for them and pursuing it rely on different mechanisms, such as nonrepresentational forms of chemical signalling. But in the end, these are just different tools that different kinds of organisms have developed to do the very same thing: to seek what is good for them and avoid what is bad for them. Kenneth Goodpaster argues that it is what we seek and avoid—i.e., the ends that these valenced conscious states or other mechanisms subserve—that should be thought of as constituting a thing's welfare. He says,

Biologically, it appears that sentience is an adaptive characteristic of living organisms that provides them with a better capacity to anticipate, and so avoid, threats to life. This at least suggests ... that the capacities to suffer and to enjoy are ancillary to something more important rather than tickets to considerability in their own right¹⁹.

And then, quoting Mark Lipsey:

If we view pleasure as rooted in our sensory physiology, it is not difficult to see that our neurophysiological equipment must have evolved via variation and selective retention in such a way as to record a positive signal to adaptationally satisfactory conditions and a negative signal to adaptationally unsatisfactory conditions . . . The pleasure signal is only an evolutionarily derived indicator, not the goal itself. It is the applause which signals a job well done, but not the actual completion of the job²⁰.

Goodpaster argues that it is the “job well done,” not the “applause” signalling it, that constitutes a being’s interests. If this is right, then in asking which things have interests, we should ask which things have ‘jobs’ of this sort, rather than asking whether a being signals success or failure to itself in a particular way. That is to say, we should look at which things have biological goals along with behaviour directed at achieving those goals rather than at which things use a particular tool for achieving them. Individualists claim that all individual organisms have biological goals but that ecological wholes do not, and this is the basis for their claim that ecological wholes are not morally considerable.

So what exactly are biological goals, and how do we determine which things have them? One might worry that explaining the behaviour of nonconscious living things as a matter of their seeking the achievement of certain ends invokes a kind of Aristotelian teleological understanding of the world that has been discarded by contemporary biology²¹. Contemporary biologists do not see organisms as having particular aims or purposes built into them, and so the kinds of teleological explanations offered by Aristotle are unavailable to those who wish to make their views consistent with contemporary science.

An important task for the individualist, then, is to find an understanding of what goal-direction might amount to that is consistent with contemporary biology. Luckily, philosophers of biology have been working on this very problem, as biology still has a need for teleological explanations consistent with contemporary science. The explanation of goal-direction that most prominent contemporary individualists (e.g., Gary Varner and Nicholas Agar) have relied on is the etiological account originated by Larry Wright. On this view, a thing’s biological goals are just the fulfillment of its biological functions. Biological functions are then explained this way: for F to be a [biological] function of some trait T is for it to be the case that

- (a) T is there because it does F [i.e., it was selected for because of its performance of F in the evolutionary past], and
- (b) the performance of F is a consequence of T’s being there²².

While the second clause refers to a causal connection between the trait and the achievement of the biological goal, the first clause is usually read as an evolutionary historical claim²³. So, for example, we humans have hearts because they pump blood—i.e., it is the pumping of blood by past hearts that is, via evolutionary selective mechanisms, responsible for the existence of hearts in humans today. That is to say, hearts were selected for in humans because they pump blood. Adding to this the claim that our blood gets pumped as a consequence of our having hearts, we get the claim that the function of our hearts is to pump blood, and that fulfilling this function is a biological goal of ours, and hence one of our interests.

But here is the tricky move. Since a thing has interests only if it has biological goals, and it has biological goals only if it has traits with biological functions, and

it has traits with biological functions only if its traits were selected for because they performed those functions in the evolutionary past, then anything with interests must have traits that are subject to natural selection. And which things have traits that are subject to natural selection? This is the very question at the core of the units of selection problem in the philosophy of biology. Broadly speaking, a unit of selection is the thing that has the trait that gets selected for by natural selection. This, then, is what gets us the relationship between being a unit of selection and being a thing that has interests: since something's interests are just its traits doing what they were selected to do, only things that have such traits will have interests. So to figure out which things have interests, we just need to look at which things have the traits on which natural selection operates. If ecosystems are not units of selection, they cannot have interests. Individualists have argued that ecosystems are not in fact units of selection (a claim that was uncontroversial within philosophy of biology, as ecosystems do not reproduce and hence have no inclusive fitness that could be differentially affected by the possession of a trait), and so concluded that ecosystems do not have interests.

CRITICISMS OF THE ARGUMENT

The first complication comes from developments in work on the units of selection problem in the philosophy of biology. When this argument was first made by individualists, and for a good long while thereafter, there were epic battles going on in philosophy of biology over the units of selection problem, most publicly between Richard Dawkins and Stephen Jay Gould²⁴. While Darwin had assumed that individual organisms are the unit of selection, Gould argued that groups are as well, whereas Dawkins argued that only genes are. Thanks to work by David Hull and Robert Brandon, among others, it is now clear that proponents of different positions in these debates largely had different meanings of 'units of selection' in mind. As noted by Elliot Sober and David Sloan Wilson, and also separately Elisabeth Lloyd, the literature now typically refers not to 'units of selection' but to different roles that different entities might have within the selective process²⁵. In the literature today, we get:

replicators: things that produce copies of themselves²⁶; (in Dawkins' language, active, germ-line replicators)

interactors: things that directly interact with the selective environment; and

evolvers (Lloyd's term is "manifestor-of-adaptations"; Hull's is "lineages"): things that manifest adaptations as a result of the evolutionary process²⁷

So the general view in philosophy of biology today is that to ask "Which things are units of selection?" is to ask a confused question. There are different things we might mean by 'units of selection,' and so we can at best ask which things are *replicators*, or which things are *interactors*, or which things are *evolvers*. If we were to update the individualist argument in light of this development, which of these roles should we think qualifies something as a bearer of interests? That is to say, which one of them is the thing that has the traits that get selected for by

natural selection? There is a rough consensus that only genes are replicators; so if we choose replicators, only genes will end up having interests—on this view, not even individuals would count as having a good of their own²⁸. Interactors are typically regarded as phenotypes, which would have the result that only phenotypes have a good of their own. That said, since phenotypes seem to be traits, we might ask what they are traits of. Strictly speaking, phenotypes are phenotypes of genotypes in environments, which would bring us back to only genes having interests. But we might try to ask, even more loosely, which entity is the ‘vehicle’ that ‘carries’ the phenotype (to use Dawkins’ language). Here there might be room for individuals, though Dawkins’ arguments for the ‘extended phenotype’ show that phenotypes can be traits of other organisms besides the one possessing the relevant genes, as well as traits of groups and, some have argued, even traits of ecosystems²⁹. For example, Dawkins describes the case of a snail parasite, which increases the fitness of its own genes by making the shell of its host snail thicker and thus harder to crush; likewise, many social animals exhibit behaviour at the group level that increases the fitness of their genes, and (again from Dawkins) the lakes that result from beaver dams increase the fitness of the beaver’s genes³⁰. The analysis of evolvers is complicated by debate about what to count as an adaptation, but the two main candidates in the literature seem to be species (Hull’s “lineages”) or genes (Dawkins’ “genes”)³¹. On this view, neither individuals nor ecosystems would count as bearers of interests.

None of the new categories can get the individualist quite what she wants, which is the claim that individuals, but not wholes, have interests. If we look at replicators, then only genes have interests; if we look at interactors, then individuals, groups, ecosystems, and many other kinds of things can have interests; if we look at evolvers, then either genes or species but not individuals have interests. So the defense of individualism by identifying bearers of interests with units of selection has been undermined by the turn that the units of selection debate has taken. Upon reflection, this should not be terribly shocking. Philosophers of biology are not interested in determining which things have a good of their own, and so their classifications are neither constructed in light of nor answerable to facts about what is or is not relevant to the possession of interests. Philosophers of biology are interested in arriving at the most accurate way of modeling how evolutionary processes work; accordingly, things that play a slightly different role in those models (a) ought to be classified differently and (b) ought to be classified by the role that they play.

This turn in the units of selection debate thus makes it much more difficult to use ‘unit of selection’ as a proxy for ‘bearer of interests,’ and I think there are further reasons not to want to identify these two notions in the first place. Indeed, progress on the units of selection problem might motivate us to revisit the individualist attraction to units of selection—an attraction that I think was ill-founded to begin with.

Recall that we arrived at the concern with units of selection by looking for something about an entity that could count as ‘its good.’ While we seemed to arrive at a nonarbitrary feature of organisms that we could use to give content to its interests, this feature does not seem to give us the content that we should think of as constituting the organism’s interests. The etiological account of functions, as its proponents readily admit, is mostly backward-looking: F counts as a function of T if and only if T still does F and T’s performance of F contributed to the evolutionary success of possessors of T in the past. When we use this as the basis for interest-attributions, we are essentially saying that what it is good for a thing to do *in the present* is whatever conferred a selective advantage on that thing’s ancestors *in the past*. But this will coincide with our intuitive sense of what is good for a thing only in cases where the thing’s current selective environment is largely the same as its ancestors’ environment was. In cases where behaviours that previously conferred a selective advantage now make it more difficult for an organism to survive, we are rightly reluctant to say that continuing to exhibit this behaviour is nonetheless good for the organism.

How serious a problem is this? Recall the problems we face with climate change. Most experts believe that the pace of changes in the climate are such that environmental change will happen too rapidly for adaptation by natural selection to keep up with. Hence the dire predictions about huge rates of extinction in the next century. I contend that the individualist argument above only looked plausible as an account of interests against a background in which a thing’s continuing to do what it evolved to do, what is ‘natural’ for it to do, continues to be what is good for it. In an environment where most things are pretty well adapted to their current circumstances, their ‘natural’ behaviours and the behaviours that are good for them will tend to coincide. But crucially, the background has changed and will continue to do so. Increasingly, we face a world where maladaptation is the rule rather than the exception—where an organism’s natural behaviours might well doom it in its current context. The coincidence assumed by individualism between ‘natural’ behaviours and advantageous behaviours is disappearing.

This brings us to a further criticism. As John Basl and Ron Sandler have pointed out, the backward-looking method for establishing the content of interests is also a problem for synthetic organisms—i.e., those that exist and have the features they do due to human engineering rather than evolutionary history³². Such organisms will turn out not to have interests at all, even if they behave in the same ways and with the same effects as evolved organisms. This produces what Basl and Sandler call the “symmetry puzzle”: two organisms can have perfectly identical intrinsic properties, but if one is the product of evolution and the other of human engineering, the first one will have interests and the second will not. This strains credulity: one would expect identical organisms to have the same interests. Again, in such cases, looking to the evolutionary past to determine what is good for something in the present does not appear to be a very good strategy. My assessment of the defence of individualism, then, is that while these arguments might be able to find something nonarbitrary about organisms

that can give content to their interests, the content identified is not what we should consider their interests to be.

There are further reasons for worrying about the superiority of individualism over holism. The more we learn about organisms, the more we see that many turn out to be, as a matter of both evolution and biological functioning, a lot more like wholes than like individuals. Human beings, obvious bearers of interests with which the individualists began their argument, are a case in point. Medical research increasingly studies not the human body, but the human biome. We humans, it turns out, share our bodies with approximately 10,000 species of bacteria. Bacterial cells outnumber “human” cells in our bodies by a ratio of 10 to 1³³. Their behaviour regulates our metabolism, our immune system, our moods, and much, much more. They do so much of what goes on in our bodies that at least one scientist has concluded “the model that places our genes at the root of all human development is wrong³⁴.” This is why medical researchers have gotten so interested in ecology lately: they hope to develop a much more accurate understanding of how humans work by studying humans as ecosystems rather than as individuals.

Functionally, then, organisms are not as easily individuated as one might have thought: we operate a lot more like ecosystems than like discrete individuals. However, a defender of individualism might concede this point but nonetheless insist that it is our DNA that makes us unique. After all, the way that the bacterial cells were distinguished from the human cells to arrive at the “10 to 1” ratio mentioned above was by looking at what kind of DNA each cell has. So perhaps we can individuate organisms genetically. While everything in my body might not count as “me,” at least all of the cells in my body with my DNA do. An organism, then, is whatever has its DNA; organisms with different DNA count as part of the organism’s environment rather than as part of the organism itself.

But this method of individuation won’t quite get us discrete individuals either. The distinction between organism and environment turns out to be rather messy, as do the contents of our genetic code itself. New developments in evolutionary biology have complicated our picture of how evolutionary mechanisms work. Epigenetics has shown that the relationship between genes and the environment is more complicated than we had thought. Not only can the environment alter an organism’s genes by switching genes on and off, but it can produce heritable changes that are passed on to subsequent generations (through epigenetic tags added to DNA and then replicated as cells divide)³⁵.

Finally, there is the matter of our DNA itself. Aside from the obvious point that identical twins share the same DNA (yet I hope we would be reluctant to call the two of them *one* organism), advances in microbiology have shown that the genomes of many organisms contain endogenous elements: genetic material inserted into germ-line DNA by viruses or other parasites and then reproduced in the genome of subsequent generations. What people used to call “junk DNA,” i.e., the 98% of human DNA which does not code for proteins and which does

not appear to have any epigenetic function, is now understood to contain, and might even turn out to be entirely made up of, genetic material inserted into our DNA by other creatures³⁶. If that is right, then the majority of our DNA might not be, strictly speaking, *our own*.

In sum, new developments in the biological sciences are making the line between individuals and wholes a lot less clear. What human beings *are as systems* now looks a lot more like ecosystems than like the discrete individuals we used to consider ourselves to be. Furthermore, the genetic code that one might think could mark us off as individuals is not very well distinguishable from that of our evolutionary companions. Even the distinction between our genes as replicators and the selective environment they encounter is not as clear as might have been thought: our environment produces heritable changes in the genome of an organism just as germ-line mutations do.

A hardcore individualist should find this threatening. If individuals turn out to be wholes, then individuals cannot have interests. Or, what is more likely at the moment, if the line between individuals and wholes becomes very unclear, then it will become correspondingly unclear which things have interests. On the other hand, if we could say that being a whole does not necessarily mean being something that cannot have a welfare, then there need not be any potential threat to our moral considerability. That is to say, if moral considerability did not rest on being an individual *as opposed to* a whole, then these developments in the sciences need not undermine our moral sense of what we are responsible to in the world.

I think that the preceding arguments establish two claims: first, that in order to meet the challenges to environmental policy and to our understanding of individuals, we have a strong pragmatic interest in finding a way talk about the welfare of wholes; second, that existing arguments meant to vindicate biocentric individualism pose no threat to our ability find one. However, none of this is yet to show that, or how, we might vindicate the literal attribution of interests to wholes.

POSSIBLE ALTERNATIVES

While the aforementioned project is too large a task to carry out here, it is worth considering the alternative theoretical responses to the challenge to individualism described above that might be available to ethical theorists. I briefly describe the five that seem most plausible, along with the advantages and disadvantages of each.

First, and most conservatively, we might remain individualists, but take the new, more complicated, understanding of what organisms are as a reason to adopt a more permissive conception of an 'individual.' Such an understanding might allow some ecosystems to count as individuals and hence as bearers of interests. A more radical revision of our notion of an individual might accept a kind of nominalism about individuals, asserting that the world contains nested systems

at different scales, with different degrees of cohesion and complexity, and that it is up to us to decide which of these to count as individuals for the purposes of morality. (One imagines Bentham rolling over in his grave—or, more accurately, in his display case—at this thought, since it clearly opens the door for the State to have interests that are not just an aggregation of the interests of individual citizens.) While this might address the problem of the blurry boundary between wholes and individuals, there is one important problem it does not solve, namely, how to determine what the interests of nonconscious individuals consist in. We still need a reason for thinking that whatever we consider their interests to be is not just a matter of our projecting our own interests onto them. The nice thing about the individualist argument was that it did seem to solve this problem—claims about the interests of organisms were as much mind-independent facts about the world as claims about evolutionary history (since they *were* claims about evolutionary history). Having lost that solution, the question is what we might replace it with.

Second, we might abandon the commitment to individualism and look around for some fact about wholes that could serve as the nonarbitrary basis for our claims about their interests. This is the approach that has been most popular in conservation biology, where claims about ecosystem health, ecosystem integrity, ecological sustainability, and biodiversity have remained common in spite of the individualist arguments aimed at undermining their normative status as ecosystemic interests. The outstanding challenge to this approach is to show why this basis for the attribution of interests is not arbitrary—i.e., what reason we have for thinking that these states really are what is good for the ecosystem (as opposed to good for us with regard to the ecosystem, or what we imagine we would want if we were ecosystems, etc.).

Third, we might give up the idea that for a thing to have interests, it must have something akin to a subjective point of view on the world—some aspect of it that is unique to it and that can give content to its interests. It is worth pointing out that in theories of human well-being, one's own subjective states are often not regarded as unproblematically giving content to one's interests. Perfectionists, for example, regard general facts about human nature as determining what is good for one. The capabilities approach grounds claims about human welfare in general claims about the kinds of creatures humans are. Even preference-satisfaction theorists are typically only willing to countenance preferences that meet certain objective evaluative standards as constituting one's interests. While there are some hedonists still around, I think it is fair to say that these days it is at least controversial what role subjective states play in determining the content of human interests. That said, while ethical theory has moved away from deference to an individual's subjective point of view on the world in determining individual welfare, political philosophy has moved toward it. The importance of active citizen participation in democratic decision-making, the value of worker participation in corporate governance, the understanding of political representation as properly a dialogical process—in political philosophy, the importance of the individual's own unique point of view on the world, and the value of its repre-

sentation in political decision-making, has grown. The further separation of ethical theory from political theory might be a fallout of this third approach. The challenge for this approach would be to show how it avoids the kind of instrumentalism toward others that motivated the concern to put the individual's perspective at the center of morality in the first place. That is to say, the third approach needs to show what we can point to as evidence that the interests we are attributing to other things really *are* their interests—rather than the interests of the State, say, or our own interests projected onto them.

Fourth, we might give up the kind of welfarist assumption that seems to be in the background of this whole discussion, namely that what we are responsible to in the world is only, or perhaps even most importantly, the interests of others. If it is a good thing that ecosystems be in certain kinds of states rather than others, then perhaps we need not show additionally that it is *good for* the ecosystems in question. On this view, the discussion of moral obligation that starts with moral considerability is a kind of confusion. This approach, I think, faces two challenges. First, a justificatory challenge: showing why biodiversity, for example, would be a good thing if it is of no benefit to anyone or anything. Second, an explanatory challenge: showing how non-welfare goods are to be treated when they conflict with welfare-goods—e.g., when biodiversity and welfare conflict, which ought to be considered more important and why.

Finally, we might take all of this to be an argument that we should not accept any kind of biocentrism at all. That is to say, perhaps we should not be willing to attribute interests to anything that has no subjective states. Going in this direction leaves us with a familiar set of problems, both practical (e.g., we have to formulate environmental policy by looking at what is good for conscious individuals) and philosophical (e.g., a world with rich, flourishing ecosystems but no conscious life turns out to have no value at all). That said, if coupled with an abandonment of the welfarist assumption (the assumption that all *good* must be *good for*), we might be able to consider ecological damage to be morally bad without having to construe it as a harm to the ecosystem. We would, of course, still need to explain what counts as damage and why it is a bad thing.

In any case, what I hope I have shown here is that even with all of the standard welfarist assumptions, the individualist argument against holism simply is not as good as many have thought. If one is going to be a biocentrist, then, it is not obvious that one ought to be an individualist.

NOTES

- * The author wishes to thank audiences at the Centre de Recherche en Éthique de l'Université de Montréal's "Attitudes, Values and Environment" workshop and the International Society for Environmental Ethics' annual meeting, at which earlier versions of this paper were presented. She thanks Jonathan Maskit in particular, for helpful comments on an earlier draft.
- ¹ For discussions, see Harris *et al.* (2006); Sandler (2012).
 - ² For discussion of this aspect of ecological restoration, see Hourdequin and Havlick (2011); Egan (2006); Hobbs *et al.* (2004); Higgs (2003).
 - ³ Egan and Howell (2001); Higgs (2003).
 - ⁴ Harris *et al.* (2006); Sandler (2012, pp. 47-74).
 - ⁵ For an extended discussion, see Sandler (2012). There are other criticisms of the legitimacy of historical benchmarks in ecological restoration, for example those questioning the legitimacy of picking out the state of a system during a certain historical time period as the 'natural' state of that system.
 - ⁶ See Sandler (2012) for a discussion of this issue.
 - ⁷ See, e.g., Primack (2004, p. 183).
 - ⁸ See, e.g., McLachlan *et al.* (2007); Mawdsley *et al.* (2009); Hoegh-Guldberg (2008); and Seddon (2010).
 - ⁹ See, e.g., Ricciardi and Simberloff (2009a; 2009b); Schlaepfer (2009).
 - ¹⁰ For discussions of the intuitive appeal of holism, see Nelson (2010).
 - ¹¹ Some writers (e.g., Regan 1983) refer to this as intrinsic or inherent value. How we should think about the relationship between value and considerability is not an issue I will take up here.
 - ¹² See, for example, Singer (1990); Varner (1998); Goodpaster (1978). For a criticism, see O'Neill (2001, pp. 168-170).
 - ¹³ For something to be in your interest is for it to be beneficial to you. As Regan (1981) points out, there is a difference between taking an interest in X and X's being in your interest. The question here is not which things can take an interest in something else, but rather, of which things it is correct to say that there is something that is in their interest.
 - ¹⁴ For an example of this kind of reasoning within animal ethics, see Regan (1983).
 - ¹⁵ See, for example, Shklar (1989).
 - ¹⁶ Bentham (1789).
 - ¹⁷ One finds precursors of this argument in Schweitzer (1929). It was popularized more recently by Goodpaster (1978). One also finds versions of the same reasoning in Rolston (1988), Agar (2001), Varner (1998); Callicott (1980). Agar also attributes this view to Paul Taylor. While this is the most common and influential argument for biocentrism, I do not mean to suggest that it is the only argument that has ever been given. Its influence, however, is such that most contemporary arguments for individualism resemble it quite closely.
 - ¹⁸ The mechanisms involve a chemical signaling and response function called chemotaxis. For a description, see Willard & Devreotes (2006).
 - ¹⁹ Goodpaster (1978, p. 316).
 - ²⁰ Goodpaster (1978, pp. 316-317), quoting Mark W. Lipsey, "Value Science and Developing Society," paper delivered to the Society for Religion in Higher Education, Institute on Society, Technology and Values (July 15-Aug. 4, 1973), p. 11.
 - ²¹ Sober (1986). Indeed, Taylor invokes the Aristotelian concept of the "telos" in explaining his view. See Taylor (1980, pp. 119-129); for a discussion see Hauskeller (2005).
 - ²² This definition is a modified version of the one given in Wright (1973, p. 161). See n. 24 for an explanation of the modification.
 - ²³ It is worth noting that Wright's original definition does not require that the causal history involve natural selection. Later proponents of his account, such as Millikan (1984) and Neander (1991) have added this requirement (or something quite close to it), and individualists in environmental ethics have typically followed their lead. In the case of Millikan and Neander,

the reason is that they were trying to explicate a notion of biological function in particular rather than function in general. Individualists in environmental ethics had an interest in ruling out the attribution of interests to artifacts (cars, computers, and the like), and so requiring a causal history involving natural selection seemed to accomplish this goal. For an example of this reasoning in environmental ethics and a discussion of some of the problems that result from it, see Basl and Sandler (2013).

²⁴ See, for example, Gould (1997). This debate is described in detail in Sterelny (2003).

²⁵ Sober and Wilson (1994); Lloyd (2007).

²⁶ Or more precisely, things of which copies get made. See Lloyd (2007, p. 45).

²⁷ The definition is paraphrased from the formulation in Lloyd (2001, p. 284). Lloyd credits Hull (1980) with the introduction of the term ‘evolvers.’ See also Hull (1978; 1980) for his discussion of ‘lineages.’

²⁸ The explanation of why individuals are not replicators is as follows: sexually reproducing organisms do not make copies of themselves at all; asexually reproducing organisms do not pass along changes to their offspring, as mutated genes do. For a discussion, see Agar (2001, p. 111) and Dawkins (1982).

²⁹ See Dawkins (1982). For examples of phenotypes manifest in other organisms, see Dawkins (1982, pp. 209-249). For examples of phenotypes manifest in ecosystems or other larger wholes, see Whitham *et al.* (2003). See also Dawkins (1982); Laland (2004); and Dawkins (2004) for a discussion of niche construction. Agar argues against what he considers a “too much extended phenotype” on pragmatic grounds: he claims that conventional biologists and researchers into genetic diseases typically are not interested in phenotypic effects quite so broad, and so we are justified in limiting our attribution of phenotypes to those that do not “go beyond the skin” of the organism possessing the genes that are causally responsible for them (Agar, 2001, pp. 111-114). However, not only are there other branches of science where scientists are interested in broader phenotypic effects (e.g., ecology), but also, since phenotypic attributions are being used as the basis for interest-attributions and hence moral considerability, one might think that the research interests of a certain class of scientists is not a legitimate basis for including or excluding entities from these moral domains.

³⁰ Dawkins’ example of the trematode parasite is discussed in Agar (2001, p. 111). For further examples and explanation, see Dawkins (1982; 2004).

³¹ Lloyd (2001, p. 272).

³² Basl and Sandler (2013).

³³ Turnbaugh *et al.* (2007).

³⁴ Specter (2012), quoting Martin J. Blaser, Professor of Microbiology at the New York University School of Medicine.

³⁵ Armstrong (2014).

³⁶ Holmes (2011).

REFERENCES

Agar, Nicholas, *Life's Intrinsic Value: Science, Ethics and Nature*, New York, Columbia University Press, 2001.

Armstrong, Lyle, *Epigenetics*, New York, NY, Garland Science, 2014.

Basl, John and Ronald Sandler, "Three Puzzles Regarding the Moral Status of Synthetic Organisms," in Gregory E. Kaebnick and Thomas H. Murray (eds.), *Synthetic Biology and Morality: Artificial Life and the Bounds of Nature*, Cambridge, MA, MIT Press, 2013, pp. 89-106.

Bentham, Jeremy, *The Principles of Morals and Legislation*, New York, NY, Hafner Press, [1789] 1948.

Brandon, Robert N., "The Levels of Selection," *Proceedings of the Philosophy of Science Association*, vol. 1982, no. 1, 1982, pp. 315-323.

Cahen, Harley, "Against the Moral Considerability of Ecosystems," *Environmental Ethics*, vol. 10, no. 3, 1988, pp. 195-216.

Callicott, J. Baird, "Animal Liberation: A Triangular Affair," *Environmental Ethics*, vol. 2, no. 4, 1980, pp. 311-328.

Dawkins, Richard, *The Extended Phenotype: The Gene as the Unit of Selection*, Oxford, W. H. Freeman and Company, 1982.

Dawkins, Richard, "Extended Phenotype – But Not Too Extended: A Reply to Laland, Turner and Jablonka," *Biology and Philosophy*, vol. 19, no. 3, 2004, pp. 377–396.

Egan, Dave and Evelyn A. Howell (eds.), *The Historical Ecology Handbook: A Restorationist's Guide to Reference Ecosystems*, Washington, D.C., Island Press, 2001.

Egan, Dave, "Authentic Ecological Restoration," *Ecological Restoration*, vol. 24, no. 4, 2006, pp. 223-224.

Goodpaster, Kenneth E., "On Being Morally Considerable," *Journal of Philosophy*, vol. 75, no. 6, 1978, pp. 308-325.

Gould, Stephen Jay, "Darwinian Fundamentalism," *New York Review of Books*, June 12, 1997.

Harris, James A., Richard J. Hobbs, Eric Higgs, and James Aronson, "Ecological Restoration and Global Climate Change," *Restoration Ecology*, vol. 14, no. 2, 2006, pp. 70-76.

Hauskeller, Michael, "Telos: The Revival of an Aristotelian Concept in Present Day Ethics," *Inquiry: An Interdisciplinary Journal of Philosophy*, vol. 48, no. 1, 2005, pp. 62-75.

Higgs, Eric, *Nature by Design: People, Natural Process, and Ecological Restoration*, Cambridge, MA, MIT Press, 2003.

Hobbs, Richard J., Mark A. Davis, Lawrence B. Slobodkin, Robert T. Lackey, William Halvorson and William Throop, "Restoration Ecology: The Challenge of Social Values and Expectations," *Frontiers in Ecology and the Environment*, vol. 2, no. 1, pp. 43-48.

Hoegh-Guldberg, O., L. Hughes, S. McLntyre, D. B. Lindenmayer, C. Parmesan, H. P. Possingham and C. D. Thomas, "Assisted Colonization and Rapid Climate Change," *Science*, vol. 321, no. 5887, 2008, pp. 345-346.

Holmes, Edward C., "The Evolution of Endogenous Viral Elements," *Cell Host and Microbe*, vol. 10, no. 4, 2011, pp. 368-377.

Hourdequin, Marion and David G. Havlick, "Ecological Restoration in Context: Ethics and the Naturalization of Former Military Lands," *Ethics, Place, and Environment*, vol. 14, no. 1, 2011, pp. 69-89.

Hull, David L., "A Matter of Individuality," *Philosophy of Science*, vol. 45, no. 3, 1978, pp. 335-360.

Hull, David L., "Individuality and Selection," *Annual Review of Ecology and Systematics*, vol. 11, 1980, pp. 311-332.

Laland, Kevin N., "Extending the Extended Phenotype," *Biology and Philosophy*, vol. 19, no. 3, 2004, pp. 313-325.

Leib-Mosch, C., R. Brack-Werner, T. Werner, M. Bachmann, O. Faff, V. Erfle, and R. Hehlmann, "Endogenous Retroviral Elements in Human DNA," *Cancer Research*, vol. 50, sup. 17, 1990, pp. 5636s-5642s.

Lloyd, Elisabeth A., "Units and Levels of Selection: An Anatomy of the Units of Selection Debates," in Rama S. Singh, Costas B. Krimbas, Diane B. Paul, and John Beatty (eds.), *Thinking about Evolution: Historical, Philosophical, and Political Perspectives*, Cambridge, Cambridge University Press, 2001, pp. 267-291.

Lloyd, Elisabeth A., "Units and Levels of Selection," in David Hull and Michael Ruse (eds.), *The Cambridge Companion to the Philosophy of Biology*, Cambridge, Cambridge University Press, 2007, pp. 44-65.

Mawdsley, Jonathan R., Robin O'Malley, & Dennis S. Ojima, "A Review of Climate-Change Adaptation Strategies for Wildlife Management and Biodiversity Conservation," *Conservation Biology*, vol. 23, no. 5, 2009, pp. 1080-1089.

McLachlan, Jason S., Jessica J. Hellmann, and Mark W. Schwartz, "A Framework for Debate of Assisted Migration in an Era of Climate Change," *Conservation Biology*, vol. 21, no. 2, 2007, pp. 297-302.

Millikan, Ruth Garrett, *Language, Thought, and Other Biological Categories*, Cambridge, MA, MIT Press, 1984.

Neander, Karen, "The Teleological Notion of Function," *Australasian Journal of Philosophy*, vol. 69, no. 4, 1991, pp. 454-468.

Nelson, Michael P., "Teaching Holism in Environmental Ethics," *Environmental Ethics*, vol. 32, no. 1, 2010, pp. 33-49.

O'Neill, John, "Meta-Ethics," in Dale Jamieson (ed.), *A Companion to Environmental Philosophy*, Malden, MA, Blackwell Publishers, 2001.

Primack, Richard B., *A Primer of Conservation Biology*, 3rd edition, Sunderland, MA, Sinauer Associates, 2004.

Regan, Tom, "The Nature and Possibility of an Environmental Ethic," *Environmental Ethics*, vol. 3, no. 1, 1981, pp. 19-34.

Regan, Tom, *The Case for Animal Rights*, Berkeley, University of California Press, 1983.

Ricciardi, Anthony and Daniel Simberloff, "Assisted Colonization is Not a Viable Conservation Strategy," *Trends in Ecology and Evolution*, vol. 24, n. 5, 2009a, pp. 248-253.

Ricciardi, Anthony and Daniel Simberloff, "Assisted Colonization: Good Intentions and Dubious Risk Assessment," *Trends in Ecology and Evolution*, vol. 24, no. 9, 2009b, pp. 476-477.

Rolston, Holmes, III, *Environmental Ethics*, Philadelphia, Temple University Press, 1988.

Sandler, Ronald L., *The Ethics of Species: An Introduction*, Cambridge, Cambridge University Press, 2012.

Schlaepfer, Martin A., William D. Helenbrook, Katherina B. Searing, Kevin T. Shoemaker, "Assisted Colonization: Evaluating Contrasting Management Actions (and Values) in the Face of Uncertainty," *Trends in Ecology and Evolution*, vol. 24, no. 9, 2009, pp. 471-472.

Schweitzer, Albert, *Cultural Philosophy II: Civilization and Ethics*, London, A. C. Black, 1929.

Seddon, Philip J., "From Reintroduction to Assisted Colonization: Moving along the Conservation Translocation Spectrum," *Restoration Ecology*, vol. 18, no. 6, 2010, pp. 796-802.

Shklar, Judith, "The Liberalism of Fear," in Shaun. P. Young (ed.), *Political Liberalism: Variations on a Theme*, Albany, NY, SUNY Press, 1989, pp. 149-166.

Shrader-Frechette, Kristin, "Sustainability and Environmental Ethics," in John Lemons, Laura Westra and Robert Goodland (eds.), *Ecological Sustainability and Integrity: Concepts and Approaches*, Dordrecht, Kluwer, 1998, pp. 16-30.

Singer, Peter, *Animal Liberation*, New York, Avon Books, 1990.

Sober, Elliott, "Philosophical Problems for Environmentalism," in B. G. Norton (ed.), *The Preservation of Species: The Value of Biological Diversity*, Princeton, Princeton University Press, 1986, pp. 173-194.

Sober, Elliott and David Sloan Wilson, "A Critical Review of Philosophical Work on the Units of Selection Problem," *Philosophy of Science*, vol. 61, no. 4, 1994, pp. 534-555.

Specter, Michael, "Germs Are Us," *The New Yorker*, October 22, 2012.

Sterelny, Kim, *Dawkins vs. Gould: Survival of the Fittest*, Cambridge, Icon Books, 2003.

Varner, Gary E., *In Nature's Interests? Interests, Animal Rights, and Environmental Ethics*, Oxford, Oxford University Press, 1998.

Taylor, Paul, *Respect for Nature: A Theory of Environmental Ethics*, Princeton, Princeton University Press, 1980.

Thomas G. Whitham, Thomas G., William P. Young, Gregory D. Martinsen, Catherine A. Gehring, Jennifer A. Schweitzer, Stephen M. Shuster, Gina M. Wimp, Dylan G. Fischer, Joseph K. Bailey, Richard L. Lindroth, Scott Woolbright and Cheryl R. Kuske, "Community and Ecosystem Genetics: A Consequence of the Extended Phenotype," *Ecology*, vol. 84, no. 3, 2003, pp. 559-573.

Turnbaugh, Peter J., Ruth E. Ley, Micah Hamady, Claire M. Fraser-Liggett, Rob Knight and Jeffrey I. Gordon, "The Human Microbiome Project," *Nature*, vol. 449, 2007, pp. 804-810.

Willard, Stacey S. and Peter N. Devreotes, "Signaling Pathways Mediating Chemotaxis in the Social Amoeba, *Dictyostelium discoideum*," *European Journal of Cell Biology*, vol. 85, no. 9-10, 2006, pp. 897-904.

Wright, Larry, "Functions," *Philosophical Review*, vol. 82, no. 2, 1973, pp. 139-168.

DOSSIER **MÉTAÉTHIQUE ET MÉTAMÉTAPHYSIQUE**

FABRICE CORREIA
UNIVERSITÉ DE NEUCHÂTEL

CHRISTINE TAPPOLET
UNIVERSITÉ DE MONTRÉAL

PLUS ON MONTE PLUS ON S'AMUSE : INTRODUCTION

Que dois-je faire? Qu'est-ce qui existe? Voilà deux questions au cœur de la philosophie. La première occupe l'éthique, tandis que la seconde est celle que l'on se pose en métaphysique, deux des grandes disciplines de la philosophie. Mais pourquoi en rester à ces questions somme toute un peu faciles, par trop concrètes? La méta-éthique et la métamétaphysique relèvent le défi de répondre aux questions les plus abstraites. Ces disciplines réfléchissent sur la nature des objets, sur les outils et la méthodologie des disciplines associées – l'éthique pour la méta-éthique, et la métaphysique pour la métamétaphysique, comme on l'aura deviné –, ainsi que sur le sens des questions qu'elles se posent et leur capacité à y répondre. Si l'éthique et la métaphysique ont depuis l'Antiquité été considérées comme des disciplines philosophiques « officielles », les réflexions « méta » associées constituent depuis peu des champs de recherche à part entière.

Les questions « méta » sont certainement les plus abstraites et les plus fondamentales qui soient en philosophie. Elles concernent la nature essentielle des choses. Dans le domaine de l'éthique, par exemple, la méta-éthique pose la question de savoir ce que sont les valeurs et les normes, et plus généralement les faits moraux et normatifs. S'agit-il d'entités existant indépendamment de nous, de nos croyances, conventions et attitudes, comme le voudraient les réalistes moraux? Ou bien est-ce que leur statut ontologique serait plutôt comparable à celui des objets sociaux, comme l'argent? Les questions de cet ordre, que l'on associe à l'ontologie morale, sont étroitement liées à des questions d'épistémologie morale, comme celle de savoir comment des entités morales existant indépendamment de nous pourraient faire l'objet de croyances justifiées, voire de connaissance. Elles sont aussi liées à la sémantique portant sur la nature des énoncés moraux, ainsi qu'à la psychologie morale, puisque les jugements moraux sont souvent considérés comme plus étroitement liés à la motivation que ne le seraient des jugements portant sur des états de choses non normatifs. Ces problématiques, que l'on fait en général remonter à G. E. Moore, ont bénéficié d'un regain d'intérêt ces dernières années.

La métaphysique a pour objectif premier de proposer une description à la fois générale et complète de la réalité sous ses aspects les plus importants. Quant à la métamétaphysique, elle s'interroge tout d'abord sur la nature d'une telle entreprise : qu'est-ce qu'une question métaphysique? Qu'est-ce qu'une théorie métaphysique? Quels sont les concepts propres à cette discipline? Peut-on faire de la métaphysique purement à priori, ou doit-on faire appel à des connaissances empiriques – en particulier à celles qui proviennent des sciences de la nature – pour avancer de manière satisfaisante dans le domaine? La métamétaphysique réfléchit aussi à la viabilité de la métaphysique, par exemple en se demandant si les questions qu'elle pose – ou en tout cas certaines d'entre elles – sont substantielles plutôt que seulement verbales, et si les théories qu'elle propose ont une assise épistémologique solide. Finalement, elle examine la place de la métaphysique dans le domaine du savoir (si elle en a une) : est-elle, comme plusieurs l'ont pensé, la philosophie première? Est-elle au contraire assujettie à une autre discipline philosophique, par exemple à la logique ou à la philosophie du langage? Ou encore, n'est-elle au mieux qu'une annexe des sciences de la nature, comme W. V. Quine et ses émules l'ont pensé?

Les nouveaux développements dans chacune de ces deux métadisciplines sont non seulement intéressants en tant que tels, mais ils sont annonciateurs d'échanges fructueux entre elles. La connexion entre métamétaphysique et méta-éthique se joue sur au moins deux plans : 1) Un grand nombre de thèses de méta-éthique à propos des valeurs et des normes sont au moins partiellement métaphysiques (par exemple la thèse selon laquelle les propriétés axiologiques sont objectives, « dans le monde »), et le questionnement métamétaphysique sur la viabilité de l'entreprise métaphysique s'y applique donc directement. 2) Nombre de concepts, outils et méthodes discutés en métamétaphysique sont utilisés en méta-éthique (par exemple les concepts de connexion nécessaire, de survivance, de fait, d'objectivité, ou encore les méthodes formelles de la logique modale).

Toutefois, les opportunités de dialogue entre la méta-éthique et la métamétaphysique n'ont jusqu'ici été que peu fréquentes. Considérant que ces métadisciplines ont de nombreux éléments en commun, et ont beaucoup à gagner à interagir, c'est précisément un tel dialogue que nous voulons promouvoir à travers la publication de ce dossier. Les quatre textes rassemblés ici sont issus d'un colloque sur le thème « Méta-éthique et métamétaphysique » tenu en mars 2014 à Ovronnaz, en Suisse, organisé conjointement par la SoPhA, la CUSO et les universités de Montréal, de Neuchâtel et de Genève. Ces textes proposent, chacun à leur manière, d'élargir et approfondir des questions de métamétaphysique et de méta-éthique en vue de faire dialoguer ces deux disciplines.

Dans son article, Pablo Carnino analyse la notion de fondation, telle qu'on la trouve à l'œuvre quand on dit par exemple qu'une action est bonne *en vertu* du fait qu'elle promeut le bonheur. Même si son approche est celle d'un métamétaphysicien à part entière, la thèse qu'il défend, qui fait usage de la notion d'essence, ne manquera pas d'intéresser les méta-éthiciens. En effet, ces derniers

font constamment usage du concept de fondation, par exemple lorsqu'ils parlent de la relation entre les propriétés morales et les propriétés naturelles.

La seconde contribution, d'Antoine C.-Dussault, fait justement un ample usage de la notion de fondation. Plus spécifiquement, la problématique examinée concerne la possibilité de soutenir que la valeur finale d'une chose peut être fondée sur des propriétés relationnelles ou extrinsèques. En se basant sur une analyse des valeurs en termes d'attitudes appropriées, l'auteur soutient que c'est en vertu des propriétés essentielles, mais pas nécessairement intrinsèques, d'un objet que ce dernier a de la valeur finale.

Le troisième texte de ce dossier appartient à la métamétaphysique et porte sur l'engagement ontologique d'une théorie. Thibaut Giraud y discute de la possibilité d'utiliser la sémantique formelle pour déterminer les engagements ontologiques d'une théorie, qu'ils soient nominalistes ou réalistes. Autant d'approches que l'on trouve aussi bien en métaphysique qu'en méta-éthique, évidemment.

Le quatrième article nous ramène à la méta-éthique. Il traite de la nature des exigences normatives et plus exactement sur ce qui les fonde. Contrairement au fait que le catholicisme exige que l'on assiste à la messe le dimanche, le fait qu'une personne souffre semble fonder une réelle exigence normative. Selon la thèse avancée par Bruno Guindon, les exigences normatives sont celles et seulement celles qui sont sensibles aux raisons. Bien que cela ne soit pas explicitement spécifié, l'argument semble présupposer une conception réaliste des raisons. En effet, l'auteur considère le fait qu'un être souffre comme une raison d'agir indépendante de nos croyances et de nos désirs.

SOURCES, RAISONS ET EXIGENCES

BRUNO GUINDON

DOCTORANT EN PHILOSOPHIE, UNIVERSITÉ MCGILL

RÉSUMÉ :

Il existe de nombreuses sources d'exigences. Certaines exigences sont normatives dans la mesure où elles impliquent des affirmations concernant ce que nous avons raison de croire, faire, désirer, etc. À ce titre, les exigences morales sont parmi les meilleures candidates. Si la morale exige que l'on tienne notre promesse, il semble que nous avons une raison de la tenir. Cependant, ce ne sont pas toutes les exigences qui sont normatives en ce sens. Le catholicisme exige que l'on assiste à la messe chaque dimanche. Il ne s'ensuit pas pour autant que nous avons une raison d'y assister. Pourquoi cela? Pourquoi certaines exigences sont normatives en ce sens, mais pas d'autres? En vue de répondre à cette question, je défends la conception *reasons-sensitive* des exigences normatives, selon laquelle les exigences normatives sont celles et seulement celles qui sont *sensibles* aux raisons. Par contre, cette conception n'est pas la seule disponible. Selon ce que j'appelle la conception *reasons-providing*, les exigences normatives sont celles et seulement celles qui *fournissent* des raisons. Je soutiens que cette conception-ci est vulnérable à deux objections, ce qui nous donne raison de préférer la première.

ABSTRACT:

There are many different sources of requirements. Some requirements are normative in the sense that they entail claims about what we have reasons to believe, intend, desire, and so on. Moral requirements are among the most plausible candidates. If morality requires that you keep your promises, it seems to follow that there is a reason for you to do so. However, not all requirements are normative in this sense. Catholicism requires that you attend Mass every Sunday. However, it does not follow that there is a reason for you to do so. Why is that? Why are some, but not all, requirements normative in this sense? In the aim of answering this question, I defend the reasons-sensitive view of normative requirements, according to which normative requirements are those and only those which are sensitive to reasons. However, the reasons-sensitive view is not the only account on offer, nor is it the most dominant. According to what I call the reasons-providing view, normative requirements are those and only those which constitute or provide reasons. I argue that the reasons-providing view faces two important objections, which gives us reason to prefer the reasons-sensitive view

INTRODUCTION

Il existe de nombreuses sources d'exigences. La bienséance exige que l'on mastique notre nourriture la bouche fermée, la prudence exige que l'on regarde des deux côtés avant de traverser la rue et le catholicisme exige que l'on assiste à la messe chaque dimanche. Il y a aussi des exigences légales, morales, et rationnelles : la loi exige que l'on paye l'impôt sur notre revenu, la morale exige que l'on considère le bien-être des animaux, et la rationalité exige que l'on ne croie pas simultanément deux propositions contraires¹. Cette liste n'est évidemment pas exhaustive.

Il y a un sens dans lequel toutes les exigences sont normatives. Les exigences sont des normes, et celles-ci expriment des standards de conformité. On peut échouer ou réussir à vivre à la hauteur de ces normes. Ainsi, les exigences sont normatives dans la mesure où elles ont affaire avec les normes. Mais cela est un sens faible du terme, et c'est un sens plus fort qui m'intéresse. Par « normatif », j'entends ce qui est nécessairement lié aux raisons²; la relation que j'invoque est celle de *l'implication*. Les exigences normatives impliquent des affirmations sur ce que nous avons une raison de faire, croire, désirer, ressentir, valoriser, etc. Plus généralement :

- (1) Nécessairement, si S est une source d'exigences normatives, alors, si S exige que $N \phi$ (où N représente un individu et ϕ un groupe verbal), il y a une raison pour N de ϕ -er.

Très peu de gens qui travaillent sur la nature des exigences normatives rejettent (1). De fait, je considère (1) comme une vérité analytique.

Par contre, toutes les exigences ne sont pas normatives en ce sens plus fort. Considérons, par exemple, les exigences du catholicisme. Le catholicisme exige que l'on assiste à la messe chaque dimanche. Mais il ne semble pas s'ensuivre pour autant qu'il y ait une raison pour moi d'y assister. Il est peut-être vrai que j'ai une raison d'y assister : peut-être que j'ai promis à ma grand-mère de l'accompagner. Mais si j'ai une raison d'assister à la messe chaque dimanche, cela ne semble pas découler du fait que le catholicisme l'exige. Il semble qu'on peut facilement imaginer un monde dans lequel le catholicisme exige que j'assiste à la messe chaque dimanche sans qu'il y ait de raison d'y assister. De fait, je pense que le monde actuel est un tel monde.

Cela semble être aussi le cas avec les exigences de bienséance. La bienséance exige que l'on mastique notre nourriture la bouche fermée. Il semble pourtant qu'on peut facilement concevoir un monde dans lequel on n'a aucune raison de mastiquer notre nourriture la bouche fermée, même s'il est vrai dans ce monde que la bienséance l'exige. Cela suggère que c'est une question importante que celle de savoir si les exigences provenant d'une source particulière sont normatives.

Il est tout de même plausible qu'il existe des exigences normatives. Les exigences morales sont peut-être les candidates les plus plausibles à ce titre. Considérons par exemple l'exigence morale voulant que l'on s'abstienne de torturer les animaux pour le simple plaisir. Dans ce cas, je ne peux imaginer un monde où la morale exigerait cela, mais où l'on n'aurait aucune raison de s'abstenir de le faire. Autrement dit, il semble évident que si la morale exige que l'on s'abstienne de torturer les animaux pour le plaisir, c'est qu'il y a une raison de s'en abstenir. Dans cet article, je suppose qu'il existe des exigences normatives et que les exigences morales en font partie. À partir de ce que nous pouvons apprendre sur la normativité de la morale, je vais premièrement offrir une conception qui peut expliquer pourquoi certaines sources d'exigences sont normatives en ce sens plus fort.

La conception que je propose est très intuitive : les exigences normatives sont celles et seulement celles qui sont *sensibles aux raisons*. Je nomme cette conception *reasons-sensitive*, et je l'étaierai dans cet article. Mais il est important de savoir qu'elle n'est pas la seule disponible, ni la plus dominante. Selon la conception concurrente que j'attribue à John Broome³, les exigences normatives sont celles et seulement celles qui *constituent* ou *fournissent* des raisons. Je nomme cette conception *reasons-providing*, et je montrerai qu'elle est vulnérable à deux objections importantes. Je présenterai ces objections et, par suite, j'expliquerai pourquoi nous devrions rejeter cette conception en faveur de la conception *reasons-sensitive*.

QUELQUES CLARIFICATIONS

J'aborde tout d'abord brièvement un sujet de préoccupation pour certains, concernant la façon dont j'utilise le terme « exigence ». Cela permettra de clarifier la conception que je développerai et la façon dont je comprends cette notion.

Il peut sembler bizarre ou même faux de dire, par exemple, que le catholicisme exige que l'on assiste à la messe chaque dimanche même si l'on n'est pas catholique, ou si l'on ne se soucie pas de vivre selon les normes catholiques. Sûrement, il est vrai que le catholicisme exige que les *catholiques* assistent à la messe chaque dimanche. « Mais puisque *je* ne suis pas catholique et que *je* ne me soucie pas de vivre selon les normes catholiques », pourrait-on insister, « le catholicisme ne peut pas exiger que *j* assiste à la messe chaque dimanche ».

Cette objection repose sur une des deux hypothèses suivantes, auxquelles je résiste. La première est la suivante : dans la mesure où une personne est soumise à une exigence, il doit y avoir un lien approprié entre ses désirs ou ses intérêts d'une part, et l'exigence en question de l'autre. Mais cela n'est pas le cas; du moins, ce n'est pas la façon dont je comprends la notion d'« exigence ». Selon la façon dont je comprends cette notion, il y a pour chaque source d'exigence *S* une propriété *P*. Ainsi, dire que le catholicisme exige que l'on assiste à la messe chaque dimanche, c'est dire qu'assister à la messe chaque dimanche est une condition nécessaire pour avoir la propriété du catholicisme. Autrement dit, si

vous êtes catholiques, vous assistez à la messe chaque dimanche. Cette proposition conditionnelle est vraie, même si vous n'êtes pas catholiques ou que vous ne vous souciez pas de vivre selon les normes catholiques⁴.

De la même façon, dire que la bienséance exige que l'on mâche notre nourriture la bouche fermée, c'est dire que mâcher notre nourriture la bouche fermée est une condition nécessaire pour avoir la propriété de bienséance. Il se peut que vous ne vous souciez pas de mastiquer votre nourriture la bouche fermée. En effet, vous préférez peut-être mastiquer la bouche ouverte. Mais cela ne change pas le fait que la bienséance exige que vous mastiquiez votre nourriture la bouche fermée. Comme le dit Philippa Foot : « [l]acking a connection with the agent's desires or interests, this [requirement] does not stand "unsupported and in need of support"; it requires only the backing of the rule⁵. »

La deuxième hypothèse sur laquelle l'objection peut être fondée postule que l'on ne peut être soumis à une exigence si l'on n'a pas de raison de faire ce qu'elle exige. L'idée semble être la suivante : puisque je ne suis pas catholique et que je ne me soucie pas de vivre selon les normes catholiques, je n'ai aucune raison d'assister à la messe chaque dimanche. Donc, le catholicisme ne peut pas exiger que j'assiste à la messe chaque dimanche.

Cependant, je suppose ici que l'on *peut* être soumis à une exigence même si l'on n'a aucune raison de faire ce qu'elle exige. Penser autrement consisterait à penser que toutes exigences sont normatives, ce que je rejette. Mais comme je l'ai suggéré précédemment au moyen d'exemples, je pense que cette hypothèse est plutôt invraisemblable. Dire qu'une source d'exigences *S* exige que l'on ϕ , c'est dire que ϕ -er est une condition nécessaire pour avoir la propriété *P* correspondant à la source *S*. Cela n'est pas encore normatif; du moins, pas dans le sens où cela impliquerait des affirmations concernant ce que nous avons une raison de faire, croire, désirer, ressentir, valoriser, etc.

Même si on ne m'accordait pas l'hypothèse postulant que l'on peut être soumis à une exigence bien que l'on n'ait aucune raison de faire ce qu'elle exige – alléguant que la notion d'exigence est intrinsèquement normative –, il demeurerait intéressant de considérer en quoi consiste la nature de cette normativité. On pourrait se demander pourquoi les exigences et les raisons voyagent toujours ensemble comme elles le font. Ma conception fournit une réponse à cette question; elle devrait donc avoir un intérêt même pour ceux qui nient l'hypothèse que je viens de formuler.

LA CONCEPTION REASONS-SENSITIVE

C'est une caractéristique commune à toutes les sources d'exigences de susciter la question de savoir pourquoi elles exigent ce qu'elles exigent. Considérons à nouveau la morale. Celle-ci exige que l'on considère le bien-être des animaux, particulièrement lorsque les conséquences de nos décisions peuvent les affecter. C'est une question importante que celle de savoir pourquoi la morale exige cela. Quelles considérations expliquent pourquoi la morale exige que l'on considère

le bien-être des animaux? Ce n'est qu'une conjecture, mais je pense qu'il est tout à fait plausible que la morale exige cela parce que les animaux (certains, du moins) sont doués de sensation et peuvent souffrir. Cette considération semble être une raison qui explique d'une manière plutôt simple et suffisante pourquoi la morale exige que l'on considère le bien-être des animaux.

Par contre, le fait qu'il y ait une raison expliquant pourquoi la morale exige que l'on considère le bien-être des animaux ne peut pas expliquer la différence que je postule entre les exigences morales et celles, par exemple, du catholicisme. Cela parce qu'il est très plausible que l'on puisse aussi citer pour chaque exigence catholique la raison pour laquelle elle exige ce qu'elle exige. Considérons par exemple l'exigence catholique que l'on assiste à la messe chaque dimanche. Comme pour les exigences morales, la question se pose de savoir pourquoi le catholicisme exige ce qu'il exige. Ici, je peux seulement supposer, mais je soupçonne que la réponse ferait appel au fait qu'assister à la messe du dimanche est une tradition de l'Église catholique, ou une règle dictée par un édit papal. Quelle que soit la réponse, l'important est qu'il existe une considération (ou un ensemble de considérations) expliquant pourquoi le catholicisme exige que l'on assiste à la messe chaque dimanche.

Dans les deux cas, on peut trouver *des raisons explicatives* et citer les raisons expliquant pourquoi chaque source d'exigences exige ce qu'elle exige. Donc, ce fait ne peut expliquer la différence entre les exigences morales et, par exemple, les exigences catholiques.

Telle que je la conçois, la conception *reasons-sensitive* est en mesure d'expliquer cette différence. De ce point de vue, c'est seulement dans le cas de la morale que la raison (ou l'ensemble des raisons) expliquant pourquoi elle exige ce qu'elle exige *compte aussi en faveur* de la réalisation ce qu'elle exige. Le fait que certains animaux sont doués de sensation et peuvent souffrir est une *considération qui compte en faveur* de la prise en compte leur bien-être lorsque les conséquences de nos décisions peuvent les affecter. Autrement dit, c'est aussi une *raison normative* de faire ce que l'exigence morale en question exige de faire.

Selon la conception *reasons-sensitive* que je défends, cela n'est cependant pas le cas pour le catholicisme. La considération (ou l'ensemble des considérations) expliquant pourquoi le catholicisme exige que l'on assiste à la messe chaque dimanche ne compte pas nécessairement en faveur d'y assister. Autrement dit, le fait qu'assister à la messe est une tradition de l'Église ou une règle dictée par un édit papal ne compte pas nécessairement en faveur de la décision d'assister à la messe chaque dimanche. Cela peut certainement expliquer pourquoi le catholicisme exige une telle pratique – ce qui revient à dire que cette considération est une raison explicative. Mais selon la conception *reasons-sensitive*, la considération (ou l'ensemble des considérations) qui détermine ou explique le contenu de l'exigence n'est pas nécessairement une raison normative d'y assister. Bien sûr, si une personne valorise l'obéissance stricte au Pape, ou si elle valorise le fait d'être (complètement) catholique, il se peut que le fait qu'assis-

ter à la messe du dimanche soit dicté par un édit papal soit une raison pour elle d'y assister. Mais cela ne démontrerait pas que le catholicisme est une source d'exigences normatives. Au plus, cela démontrerait que nos valeurs sont une source d'exigences normatives.

Contrairement aux exigences morales, même s'il n'y a aucune raison d'assister à la messe chaque dimanche, il est toujours le cas que le catholicisme exige que l'on y assiste. Par contre, s'il y n'avait aucune raison normative de considérer le bien-être des animaux – à supposer que ces animaux ne soient pas doués de sensation et ne puissent pas souffrir –, la morale n'exigerait pas que l'on considère leur bien-être⁶. Contrairement aux exigences du catholicisme, les exigences morales sont le genre d'exigences qui sont *sensibles* aux raisons. Si la morale exige que l'on considère le bien-être des animaux, c'est *parce qu'*il y a une raison morale conclusive d'envisager leur bien-être.

Bien sûr, cela doit être nuancé. Les exigences morales ne sont pas sensibles à *toutes* sortes de raisons. Il y a des considérations qui, même si elles peuvent compter en faveur de faire ce que la morale exige, ne jouent aucun rôle dans la détermination du contenu des exigences morales. Supposons qu'un milliardaire excentrique me donne un million de dollars pour favoriser le bien-être des animaux. Sûrement, cela est une raison pour le faire. Pourtant, ce fait n'explique pas, même pas en partie, pourquoi la morale exige que je considère le bien-être des animaux. Cette considération en est une d'intérêt personnel. Elle jouerait un rôle dans la détermination du contenu des exigences *prudentielles*, mais elle ne détermine pas ce que *la morale* exige. Les exigences morales sont sensibles aux raisons particulièrement *morales*⁷. Si la morale exige que l'on considère le bien-être des animaux, c'est parce qu'il y a une raison *morale* conclusive pour qu'on le fasse. Plus généralement :

(2) Si la morale exige que $N \phi$, c'est *parce qu'*il y a une raison morale et conclusive pour N de ϕ -er.

J'estime que (2) est très intuitif. Quand nous disons qu'il y a une exigence morale d'éviter un certain comportement, nous entendons que l'action est interdite par la morale. Autrement dit, l'action est immorale, auquel cas il doit exister une caractéristique qui la rend ainsi. Dans un premier temps, cette caractéristique explique pourquoi l'action en question est immorale; c'est donc une raison explicative. Pourtant, c'est aussi une raison normative; c'est une considération qui compte en faveur de l'abstention d'agir de la sorte. Dans un deuxième temps, cette raison morale pour N de ϕ -er est *conclusive*; c'est-à-dire que pour n'importe quelle autre action y , la raison pour N de ϕ -er est plus forte que toutes raisons morales pour N de ψ -er.

Je pense que (2) se généralise encore plus. Autrement dit, à partir de (2), nous pouvons obtenir la conception *reasons-sensitive* :

(RS) Nécessairement, si S est une source d'exigences normatives, alors, si S exige que $N\phi$, c'est parce qu'il y a une raison conclusive en provenance de S pour N de ϕ -er.

Bien sûr, cette inférence demande que nous supposions qu'il existe une explication unifiée des exigences normatives. En fait, elle suppose que la normativité de toutes les sources d'exigences normatives soit expliquée de la même manière que l'on explique la normativité des exigences morales. Cela est une hypothèse substantielle, mais je pense qu'elle est vraie. Au moins, je pense qu'elle devrait être l'hypothèse par défaut. Je ne vois aucune raison de multiplier les explications de la normativité des sources d'exigences au-delà de ce qui est nécessaire.

À partir de (RS), (1) suit. Supposons que S soit une source d'exigences normatives, et que S exige que $N\phi$. Selon (RS), S exigerait que $N\phi$, parce qu'il y aurait une raison conclusive en provenance de S pour N de ϕ -er. Il s'ensuivrait donc que si S exige que $N\phi$, il y aurait alors une raison pour N de ϕ -er. À cet égard, la conception *reasons-sensitive* des exigences normatives est claire et plausible.

LA CONCEPTION REASONS-PROVIDING

Selon moi, la conception *reasons-sensitive* est vraie, mais comme je l'ai dit précédemment, elle n'est pas la seule disponible, ni même la plus dominante. D'après ce que j'appelle la conception *reasons-providing*, que j'attribue à Broome :

(RP) Nécessairement, si S est une source d'exigences normatives, alors, si S exige que $N\phi$, le fait que S exige que $N\phi$ est une raison pour N de ϕ -er⁸.

De ce point de vue, les exigences normatives sont *elles-mêmes* des raisons. Si elle est vraie, cette conception explique la normativité des exigences normatives. Si les exigences normatives sont elles-mêmes des raisons, (1) découle de (RP); cela est évident. Supposons que S soit une source d'exigences normatives et que S exige que $N\phi$. De (RP), il s'ensuivrait que le fait que S exige que $N\phi$ serait une raison pour N de ϕ -er. Par instanciation existentielle, il y aurait une raison pour N de ϕ -er. Si S est une source d'exigences normatives, et que S exige que $N\phi$, il y a alors une raison pour N de ϕ -er.

Deux objections

Il est important, par contre, de voir la façon dont la conception *reasons-providing* est vulnérable à deux objections majeures, et que cela nous donne raison de préférer la conception *reasons-sensitive*. Premièrement, la conception *reasons-providing* génère un vide explicatif. La motivation de Broome en acceptant (RP) est fondée sur sa compréhension de la notion d'une exigence normative. Selon Broome, les exigences normatives sont celles (et seulement celles) qui aident à déterminer ce que nous devrions faire⁹. Supposons que la morale exige que l'on ϕ . Selon Broome, cette considération joue un rôle qui aide à déterminer si nous devrions ϕ -er. Mais maintenant, considérons comment Broome définit les « rai-

sons normatives ». Selon lui, elles font partie des explications des faits concernant ce que nous devrions faire. Plus précisément :

A consideration or a feature of a situation F is a pro tanto reason for A to ϕ iff F plays a “for- ϕ ” role in a weighing explanation of why A ought to ϕ , or in a weighing explanation of why A ought not to ϕ , or in a weighing explanation of why it is not the case that A ought to ϕ and not the case that A ought not to ϕ ¹⁰.

Mais si une considération aide à déterminer si nous devrions ϕ -er, elle fera certainement aussi partie de l’explication à savoir pourquoi nous devrions ϕ -er (ou pourquoi nous ne devrions pas ϕ -er, ou pourquoi il n’est pas le cas que nous devrions ϕ -er, et pas le cas que nous ne devrions pas ϕ -er). Selon la conception broomienne des exigences normatives, jointe à la conception broomienne des raisons, il s’ensuit que les exigences normatives sont des raisons. C’est clairement la façon dont Broome comprend la normativité des exigences morales. Selon lui :

(3) Si la morale exige que $N\phi$, alors le fait que la morale exige que $N\phi$ est une raison pour N de ϕ -er¹¹.

Toutefois, le problème est qu’il demeure important de savoir pourquoi les exigences morales, mais pas, par exemple, les exigences du catholicisme, aident à déterminer ce que nous devrions faire. Dire que les exigences morales aident à déterminer ce que nous devrions faire parce qu’elles sont elles-mêmes des raisons suscite la question suivante : pourquoi les exigences morales sont elles-mêmes des raisons, alors que d’autres exigences, par exemple, celles du catholicisme, n’en sont pas ?

Broome est conscient que cette question rôde aux alentours. Malheureusement, il ne lui offre aucune réponse. Alors que Broome *croit* (3), il admet tout de même qu’il est très difficile d’expliquer de manière convaincante pourquoi elle est vraie¹². En effet, il avoue que souvent il est difficile d’expliquer pourquoi certaines exigences sont normatives; c’est-à-dire, pourquoi elles sont des raisons. Mais à moins qu’il puisse fournir une réponse plausible à cette autre question, la conception *reasons-providing* laisse un vide explicatif.

Il est tout de même possible d’insister sur le fait que la conception *reasons-sensitive* n’est pas vraiment en meilleure position relativement à la conception *reasons-providing*. Tout comme la conception *reasons-providing*, qui laisse inexpliquée la question de savoir pourquoi certaines exigences sont des raisons, la conception *reasons-sensitive* laisse inexpliquée la question de savoir pourquoi les considérations qui déterminent le contenu des exigences normatives sont des raisons. On peut dire que les deux conceptions laissent des vides explicatifs.

Il est vrai que la conception *reasons-sensitive* n'explique pas pourquoi les considérations qui déterminent le contenu des exigences normatives sont des raisons. Cependant, cela ne génère aucun vide explicatif en elle, parce que ce n'est tout simplement pas son objectif de l'expliquer. Son objectif principal est plutôt d'expliquer pourquoi certaines – mais pas toutes – sources d'exigences sont normatives dans le sens souligné en (1), et elle l'atteint en offrant (RS). Expliquer pourquoi certaines considérations sont des raisons normatives est l'objectif des conceptions des raisons normatives; cela ne concerne pas la conception *reasons-sensitive*.

Certes, c'est aussi l'objectif principal de la conception *reasons-providing* d'expliquer pourquoi certaines – mais pas toutes – sources d'exigences sont normatives dans le sens souligné en (1). Par contre, puisque cette conception atteint son objectif en *identifiant* les exigences normatives aux raisons, elle est aussi, en partie, une conception des raisons normatives. Pour cette raison, la conception *reasons-providing* doit répondre à la question : pourquoi certaines exigences – mais pas toutes – sont des raisons normatives?

Finalement, il est possible que nous devions faire un choix. D'une part, nous pourrions accepter que les considérations qui déterminent le contenu des exigences normatives sont des raisons, et qu'il n'y ait rien de plus à dire à ce sujet. D'autre part, nous pourrions accepter que la considération que *S* exige de ϕ -er est *elle-même* une raison pour nous de ϕ -er, et qu'il n'y ait rien de plus à dire à ce sujet. Selon moi, cette affirmation est beaucoup plus difficile à accepter, et ma seconde objection explique pourquoi.

Il me paraît regrettable que la conception *reasons-providing* ne puisse expliquer pourquoi certaines exigences aident à déterminer ce que nous devrions faire, mais pas d'autres. Mais même si elle pouvait offrir une explication, je ne crois pas qu'il soit plausible que les exigences normatives soient elles-mêmes des raisons, du moins pas généralement.

Rappelons (RP). Il indique une condition suffisante :

(RP) Nécessairement, si *S* est une source d'exigences normatives, alors, si *S* exige que $N\phi$, le fait que *S* exige que $N\phi$ est une raison pour *N* de ϕ -er.

(RP) est logiquement équivalent à (4) :

(4) Nécessairement, si *S* est une source d'exigences normatives et que *S* exige que $N\phi$, alors le fait que *S* exige que $N\phi$ est une raison pour *N* de ϕ -er.

Selon (4), il suffit, pour que le fait que *S* exige que $N\phi$ soit une raison pour *N* de ϕ -er, que *S* l'exige. C'est-à-dire que c'est une condition nécessaire pour chaque exigence normative que le fait que *S* exige que $N\phi$ soit une raison pour *N* de ϕ -er. Cette condition nécessaire me paraît invraisemblable.

Considérons les données probantes, que plusieurs croient constituer une source d'exigences normatives. Les données probantes exigent plusieurs choses. Par exemple : que nous croyions en la théorie de l'évolution. Selon la conception *reasons-providing*, si les données probantes sont une source d'exigences normatives, alors le fait que les données probantes exigent que nous croyions en la théorie de l'évolution est nécessairement une raison de croire que la théorie est vraie. Mais cela n'est tout simplement pas le cas. Au contraire, les données probantes exigent que nous croyions en la théorie de l'évolution parce qu'il y a des raisons préalables et conclusives pour croire que cette théorie est vraie. Certaines de ces raisons reposent, par exemple, sur les faits concernant des traces fossiles.

Considérons également la prudence, qui exige que l'on regarde des deux côtés avant de traverser la rue. Selon la conception *reasons-providing*, le fait que la prudence exige cela est nécessairement une raison de le faire. Pourtant, ce n'est pas le cas. Comme la proposition selon laquelle les données probantes exigent que nous croyions en la théorie de l'évolution, la proposition selon laquelle la prudence exige que nous regardions les deux côtés avant de traverser la rue exprime un jugement global. Cette exigence repose sur les faits concernant le risque de traverser la rue aveuglément. C'est ce genre de faits de premier ordre qui donne lieu à l'exigence prudentielle et qui compte en faveur de la réalisation de ce qu'elle exige. Le simple fait que la prudence exige que l'on regarde des deux côtés avant de traverser la rue n'est pas une raison de le faire.

À cet égard, il y a une similitude importante entre les exigences normatives et le jugement considérant une action comme immorale. Si une action est immorale, il existe certainement des caractéristiques de l'action qui expliquent pourquoi. Ces caractéristiques sont des raisons qui comptent en faveur de l'abstention de faire l'action. Selon moi, c'est une grave erreur de croire qu'en plus des caractéristiques qui rendent l'action immorale, il y aurait une autre raison de ne pas faire l'action : c'est-à-dire, qu'elle serait immorale.

Je dois avouer que cette affirmation repose sur une thèse substantielle¹³, c'est-à-dire sur l'idée selon laquelle chaque raison normative doit être en mesure d'influencer la balance des raisons. Cela veut dire que si le fait qu'une action est immorale constitue une raison normative de ne pas agir ainsi, cette raison doit être une *autre* raison au-delà de celles qui rendent l'action immorale. Il s'agit d'une thèse controversée. Ulrike Heuer, par exemple, fait valoir qu'il est faux qu'une raison normative doit nécessairement influencer la balance des raisons¹⁴. Elle admet que le fait qu'une action soit immorale n'affecte pas la balance des raisons, mais elle soutient que le fait qu'une action soit immorale constitue tout de même une raison de ne pas agir ainsi. Son raisonnement suit un *reductio ad absurdum* : selon Heuer, il est toujours possible de spécifier les caractéristiques d'une action qui la rendent immorale. Selon ceux qui adoptent une conception *buck-passing* de la propriété d'immoralité, ce sont ces caractéristiques de premier ordre qui fournissent ou constituent les raisons normatives, et non le fait que l'action soit immorale. Par contre, plusieurs autres propriétés qui fournissent ou constituent des raisons normatives peuvent être spécifiées davantage ainsi. Heuer

cite, par exemple, le bon goût d'un mets. Le fait qu'un certain mets a bon goût constitue certainement une raison de le commander au restaurant. Il est pourtant possible de spécifier les caractéristiques du mets qui lui donnent bon goût. Il devrait donc s'ensuivre que le fait qu'un mets a bon goût ne constitue pas une raison de le commander au restaurant après tout, mais que ce sont plutôt les caractéristiques de premier ordre qui y donnent bon goût qui constituent telles raisons. Mais il est absolument clair que le fait qu'un mets a bon goût est une raison de le commander au restaurant (même si ce n'est pas nécessairement une raison conclusive). Il est donc faux qu'une propriété qui peut être spécifiée ne peut fournir une raison normative.

Je ne suis pas convaincu du raisonnement de Heuer, en grande partie parce que je doute de sa thèse selon laquelle ce qui explique pourquoi la propriété d'immoralité d'une action ne constitue pas une autre raison normative de ne pas faire l'action soit le fait que cette propriété peut être davantage spécifiée. J'éprouve en outre beaucoup de difficulté à comprendre comment une considération pourrait compter en faveur de la réalisation d'une action sans qu'elle puisse influencer la balance des raisons : les raisons ont un certain poids, ce qui affecte nécessairement la balance des raisons.

De toute façon, même s'il est faux de dire qu'une raison doit nécessairement affecter la balance des raisons, les exemples des données probantes et de la prudence que j'ai offerts ci-dessus constituent toujours un problème pour la conception *reasons-providing*. Selon cette conception, le fait que les données probantes exigent que l'on croie en la théorie de l'évolution est une raison d'y croire. Supposons que je *doive* croire que la théorie de l'évolution est vraie. Selon la conception *reasons-providing*, ce qui explique ce fait est, en partie, le fait que les données probantes l'exigent. Mais cela n'est pas une explication : cela semble plutôt être une autre façon de dire la même chose. Ce qui explique le fait que je *doive* croire que la théorie de l'évolution est vraie est le fait qu'il existe des traces fossiles qui prouvent la véracité de la théorie. Ce fait en est un de premier ordre, et c'est lui qui explique, en partie, pourquoi les données probantes exigent que nous croyions en la théorie de l'évolution.

De manière similaire, selon la conception *reasons-providing*, le fait que la prudence exige que l'on regarde des deux côtés avant de traverser la rue est une raison de le faire. Supposons donc que je *doive* regarder des deux côtés avant de traverser la rue. Selon cette conception, ce qui explique le fait que je *doive* le faire est, en partie, le fait que la prudence l'exige. Encore une fois, cela n'est pas une explication : cela semble être une autre façon de dire la même chose. Ce qui explique le fait que je *doive* regarder des deux côtés avant de traverser la rue est, en partie, le fait qu'il serait dangereux ne pas le faire. Ce fait, comme le fait qu'il existe des traces fossiles prouvant la véracité de la théorie de l'évolution, en est un de premier ordre, et c'est lui qui explique, en partie, pourquoi la prudence exige que l'on regarde des deux côtés de la rue avant de traverser.

CONCLUSION

Je conclus que même si les partisans de la conception *reasons-providing* pouvaient expliquer pourquoi certaines exigences, mais pas toutes, constituent ou fournissent des raisons, la thèse selon laquelle les exigences normatives sont nécessairement des raisons est contre-intuitive. Il est plus intuitif de penser que les raisons sont des considérations de premier ordre et que ce sont elles qui expliquent pourquoi les exigences normatives exigent ce qu'elles exigent. La conception *reasons-sensitive* accueille cette intuition et, pour cette raison, elle devrait être préférée à sa concurrente.

NOTES

- ¹ Je présume ici une logique classique. Que l'on ne tienne pas deux croyances contraires n'est pas nécessairement une exigence rationnelle dans des contextes de logiques paraconsistantes. J'ignorerai cette complication dans ce qui suit.
- ² Par « raisons », j'entends « raisons normatives »; c'est-à-dire des considérations qui comptent en faveur de la réalisation ou de l'abstention d'une action.
- ³ Voir Broome, John, *Rationality Through Reasoning*, Oxford, Wiley Blackwell, 2013; Broome, John, "Is Rationality Normative?," *Disputatio*, vol. 2, no. 23, 2007, pp. 161-178; Broome, John, "Requirements", in Rønnow-Rasmussen, Toni, Björn Petersson, Jonas Josefsson et Dan Egonsson (dir.), *Hommage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, Lund, Lund University Press, 2007, pp. 1-41; Broome, John, "Have We Reason to do as Rationality Requires?," *Journal of Ethics and Social Philosophy*, vol. 1, 2005, pp. 1-8; Broome, John, "Does Rationality Give Us Reasons?," *Philosophical Issues*, vol. 15, no. 1, 2005, pp. 312-337.
- ⁴ Il y a ici une complication mineure. Comme bien d'autres propriétés, le catholicisme est une propriété qui admet des degrés. Autrement dit, on peut être plus ou moins catholique. Ainsi, quand je dis qu'assister à la messe chaque dimanche est une condition nécessaire pour avoir la propriété du catholicisme, je veux bien sûr dire que c'est une condition nécessaire pour être *pleinement* ou *entièrement* catholique. On peut être catholique au degré *n* même si on ne parvient pas à assister à la messe chaque dimanche.
- ⁵ Foot, Philippa, "Morality as a System of Hypothetical Imperatives," *The Philosophical Review*, vol. 81, no. 3, 1972, p. 309. Foot utilise *should* au lieu de *requirement*, mais il est clair dans ce contexte qu'elle comprend *should* dans un sens non normatif.
- ⁶ C'est sûrement une caricature. On peut plausiblement affirmer qu'il existe d'autres considérations – par exemple, qui ne seraient pas fondées sur des considérations de bien-être – qui comptent en faveur de la prise en compte du bien-être des animaux. Je ne nie pas cela. Mon point est simplement que s'il n'y avait *aucune* raison d'envisager le bien-être des animaux, la morale ne l'exigerait pas.
- ⁷ Je dois admettre qu'il est difficile de dire d'une manière non circulaire ce qui fait qu'une raison est une raison morale et non, par exemple, une raison prudentielle. Il me semble intuitif de dire, un peu à la manière de Broome, qu'une raison morale est une considération qui joue un certain rôle dans l'explication du fait que la morale exige ce qu'elle exige. Par contre, quelqu'un pourrait aussi dire que, selon la conception *reasons-sensitive*, ce qui fait qu'une exigence est une exigence morale est le fait qu'elle est requise par l'ensemble des raisons morales. Ici, il semble bien y avoir un cercle; je remercie un examinateur anonyme chez *Les Ateliers de l'éthique/The Ethics Forum* sur ce point. Par contre, s'il y a un cercle ici, celui-ci n'est pas vicieux. Selon la conception *reasons-sensitive*, ce qui fait qu'une exigence est une exigence morale (et non une exigence prudentielle, par exemple) est le fait que cette exigence est une condition nécessaire pour avoir la propriété d'être morale. Il n'est donc pas nécessaire de faire appel aux « raisons morales » pour expliquer ce qui fait qu'une exigence est une exigence morale et non, par exemple, une exigence prudentielle. Il suffit de faire appel à la propriété qui correspond à la source d'exigences.
- ⁸ Broome, "Is Rationality Normative?," *op. cit.*, p. 162. Broome appelle (RP), ou quelque chose de semblable à (RP), *medium normativity*, qu'il formule plus précisément comme une thèse à propos de la normativité des exigences rationnelles. De plus, Broome surnomme (1), ou quelque chose de semblable à (1), *weak normativity*. (1) et (RP) sont plus générales, dans la mesure où elles ont affaire avec les exigences conçues généralement, plutôt qu'avec les exigences particulièrement rationnelles.
- ⁹ Broome, *Rationality Through Reasoning*, *op. cit.*, p. 126; Broome, "Requirements", *op. cit.*, p. 3.
- ¹⁰ Broome, *Rationality Through Reasoning*, *op. cit.*, p. 53.
- ¹¹ Broome, "Is Rationality Normative?," *op. cit.*, p. 165. Voir aussi Broome, *Rationality Through Reasoning*, *op. cit.*, p. 126.

¹² Broome, “Is Rationality Normative?”, *op. cit.*, p. 165.

¹³ Je remercie un examinateur anonyme chez *Les Ateliers de L'éthique/The Ethics Forum* pour m'avoir suggéré de développer ce point.

¹⁴ Voir Hueur, Ulrike, “Wrongness and Reasons,” *Ethical Theory and Moral Practice*, vol. 13, no.2, 2010, pp. 137-152.

FITTING-ATTITUDE ANALYSES AND THE RELATION BETWEEN FINAL AND INTRINSIC VALUE*

ANTOINE C. DUSSAULT

DOCTORANT EN PHILOSOPHIE, UNIVERSITÉ DE MONTRÉAL

ABSTRACT:

This paper examines the debate as to whether something can have final value in virtue of its relational (i.e., non-intrinsic) properties, or, more briefly put, whether final value must be intrinsic. The paper adopts the perspective of the fitting-attitude analysis (FA analysis) of value, and argues that from this perspective, there is no ground for the requirement that things may have final value only in virtue of their intrinsic properties, but that there might be some grounds for the alternate requirement that final value be grounded only in the essential properties of their bearers. First, the paper introduces the key elements of the FA analysis, and sets aside an obvious but unimportant way in which this analysis makes all final values relational. Second, it discusses some classical counterexamples to the view that final value must be intrinsic. Third, it discusses the relation between final, contributive, and signatory value. Fourth, it examines Zimmerman's defense of the requirement that final value must be intrinsic on the grounds that final value cannot be derivative. And finally, it explores the alternative requirement that something may have final value in virtue of its essential properties.

RÉSUMÉ :

Cet article examine, selon la perspective de l'analyse de la valeur en termes d'attitudes appropriées (la AAA), le débat concernant la possibilité qu'une chose ait de la valeur finale en vertu de ses propriétés relationnelles, ou en d'autres termes, la question de savoir si toute valeur finale doit être intrinsèque. La thèse défendue par l'article est que, selon la perspective de la AAA, il n'y a aucune raison convaincante d'adopter l'exigence selon laquelle une chose ne pourrait avoir de la valeur finale qu'en vertu de ses propriétés intrinsèques, mais il semble y avoir des bases intuitives pour adopter l'exigence alternative selon laquelle la valeur finale devrait être fondée sur des propriétés essentielles de ses porteurs. L'article présente d'abord les éléments clés de la AAA et met à l'écart une manière non pertinente selon laquelle celle-ci rend toute valeur finale relationnelle. Ensuite, l'article passe en revue quelques contre-exemples classiques à la thèse selon laquelle toute valeur finale serait nécessairement intrinsèque. Troisièmement, l'article discute de la relation entre valeurs finales, contributives et signatives. Quatrièmement, il examine la défense de l'exigence selon laquelle toute valeur finale devrait être aussi intrinsèque élaborée par Zimmerman. Et finalement, l'article explore l'exigence alternative selon laquelle une chose ne pourrait avoir de valeur finale qu'en vertu de ses propriétés essentielles.

INTRODUCTION

This paper examines the resources that fitting-attitude analyses of value (hereinafter ‘FA analyses of value’) offer for dealing with a key issue in formal axiology: whether something may have *final* value in virtue of its *relational* or *extrinsic* properties¹. In less technical terms, this issue concerns whether a thing can be valuable *for its own sake*, or *as an end*, by virtue of the relations it entertains with other things. This issue has its classical origin in G. E. Moore’s interpretation of his isolation test, according to which “[i]n order to arrive at a correct decision on the first part of this question [i.e., the question “What things have intrinsic value?”], it is necessary to consider what things are such that, if they existed by themselves, in absolute isolation, we should yet judge their existence to be good” (Moore 1988, p. 187). Moore’s idea is that, in such an isolation situation, a thing would retain all and only its *intrinsic* properties—that is, its *non-relational* properties, those that are *internal* to the thing². Moore then goes on to present the isolation test as a way of demarcating what is valuable *as an end* from what is a mere *means* to value:

By employing this method, we shall guard against two errors, which seem to have been the chief causes which have vitiated previous conclusions on the subject. The first of these is (1) that which consists in supposing that what seems absolutely necessary here and now, for the existence of anything good—what we cannot do without—is therefore good in itself. If we isolate such things, which are mere *means* to good, and suppose a world in which they alone, and nothing but they, existed, their *intrinsic worthlessness* becomes apparent (Moore 1988, p. 187, italics mine)³.

Moore’s rationale seems to be that since the relations that a thing entertains with other things would be absent under the isolation test, the value that a thing possesses *for its own sake* (as an end) must therefore be one that it has in virtue of its intrinsic properties. Hence, Moore seems to take the concepts of *final* and *intrinsic* value to be equivalent; consequently, in his view, a thing may only be valuable as an end in virtue of properties that are intrinsic to it. I will call this thesis *Moore’s intrinsicity principle*.

Moore’s intrinsicity principle has recently become the focus of a debate opposing, on the one hand, Korsgaard (1983), O’Neill (1992), Kagan (1998), and Rabinowicz and Rønnow-Rasmussen (2003; 1999), all of whom have raised serious objections against the principle, and on the other, Zimmerman (2001a; 2001b, Chap. 3), who has attempted an indirect defense of the principle by means of an argument about bearers of value⁴. This paper examines this debate from the perspective of the fitting-attitude analysis of value, which analyses the notion of value in terms of attitudes that it would be “fitting” for agents to adopt towards objects and situations of the world⁵. The thesis of the paper is that, from the perspective of the FA analysis of value, Moore’s intrinsicity principle is implausible, but that “the *essentiality* principle,” a distinct but not unrelated principle, may nevertheless be valid. Section 2 introduces the key elements of FA analy-

ses and sets aside an obvious but unimportant way in which these analyses make all final values relational. Section 3 presents classical cases put forward by Korsgaard, Kagan, O'Neill, and Rabinowicz and Rønnow-Rasmussen as counterexamples to Moore's intrinsicity principle. Section 4 discusses the relation between final value and some alleged kinds of non-instrumental extrinsic values (mainly contributive and signatory values). Section 5 examines Zimmerman's defense of Moore's intrinsicity principle through a reductionist manoeuvre he attempts along with Olson (2003). Lastly, section 6 proposes and illustrates the *prima facie* plausibility of the *essentiality* principle as an alternative to Moore's principle.

FA ANALYSES AND THE CONSTITUTION RELATION

FA analyses of value propose that to be valuable is to be a fitting object of an approving attitude (a *pro-attitude*, for short). According to such analyses, to say that something is valuable amounts to saying that it is such that it would be fitting to favour it, the notion of fittingness being generally understood in deontic terms, and the notion of favouring being usually interpreted in terms of attitudes, whence the name "fitting-attitude analyses." As its advocates typically emphasize, FA analyses thus include two main components: an *attitudinal* component (usually sentiments or emotions felt towards the object), and a *normative* component (the *fittingness* of the attitude)⁶. The attitudinal component in FA analyses portrays value concepts as concepts that, like colors and sounds, cannot be characterized independently of humans' subjective responses. This makes value concepts response-dependent, and so gives FA analyses some affinity with emotivist and other non-cognitivist metaethical theories which spell out evaluative judgments in terms of the sentiments and attitudes that evaluators *actually* experience. FA analyses dissociate themselves from these views by also including the idea of fittingness, thereby putting some distance between what evaluators *actually* feel and what is fitting to feel, that is, what they *ought* to feel towards the objects they contemplate⁷. Though FA analyses can be compatible with strong forms of non-cognitivism (depending on how the concept of fittingness is spelled out), most current versions of FA analyses conceive the fittingness of attitudes as grounded in the natural properties of the objects to which they respond. Their normative component thus grants values some independence from the evaluators' responses, and so gives FA analyses some affinity with realist metaethical theories⁸.

Three main advantages have made FA analyses appealing to many metaethicists. The first one is the simple fact that this approach actually provides an analysis of value concepts. This is not negligible: after Moore's well-known verdict that goodness is an unanalyzable concept, many value theorists simply gave up on the project of formulating an analysis of evaluative concepts. Moreover, FA analyses give such an account in a way that may relieve value properties of the air of ontological *mysteriousness* often attached to them (cf. Mackie 1977). Our sentiments and attitudes are familiar features of the world, and insofar as their fittingness can be grounded in the natural properties of the objects to which they respond, FA analyses may succeed providing an account of value that fits into a

naturalistic understanding of the world⁹. A second advantage of FA analyses is their relative metaethical neutrality. Although FA analyses exclude metaethical views that reject one or the other of their two defining components, the variety of ways in which these components can be construed leaves room for a plurality of metaethical interpretations, ranging from more realist-leaning to more anti-realist-bound ones. Such flexibility is welcome, given that realist and antirealist metaethical theories both have their strengths and shortcomings, while the debate between them is unlikely to be settled soon. And a last advantage of FA analyses is that they are also flexible at the level of axiology. FA analyses can easily accommodate the various and manifold nuances of our evaluative practices, as it allows that there may be as many distinctive ways to value something as there are shades of attitudes that an evaluator could adopt. For instance, to admire something is not equivalent to respecting it, to desire something is not the same as to be awe-inspired by it. And same goes for disvalues: for instance, to be angered by something is not equivalent to feeling contempt for it. Moreover, FA analyses are also axiologically flexible as to which kinds of thing can be proper bearers of values. Some attitudes, like desiring and hoping, seem more paradigmatically directed towards states of affairs, while others, like respecting and caring, seem more appropriate towards objects like persons or other kinds of entities. Thus, FA analyses have strong *prima facie* plausibility¹⁰. This *prima facie* plausibility suggests, in turn, that FA analyses provide a suitable perspective from which to examine the plausibility of Moore's intrinsicity principle.

It may be judicious to begin this examination by considering a very broad and general way in which FA analyses may seem to contradict Moore's principle. Given the significant role that FA analyses assign to the attitudes of evaluators in the constitution of values, it may indeed seem that such analyses unavoidably make *all* properties that have some relevance to values relational *from the outset*. This would have the consequence that *all* values are relational, running radically counter to Moore's intrinsicity principle, but in a way that seems suspiciously beside the point. As Rabinowicz and Rønnow-Rasmussen note:

One might wonder, of course, whether the claim that some values are intrinsic is compatible with the FA style of value analysis. If an object is valuable only insofar as there are reasons to have pro-attitudes toward it, then, one might argue, all value is relational by its very nature: the value of an object consists in the existence of a deontic relation between that object and the potential attitude holders. This would suggest that no value is intrinsic, since all value is relational. [...] In one sense, any value on the FA view seems to be a relational property, since it requires external attitudes toward the value bearer. But at the same time, if a given value-property of an object is grounded in its internal features, then that property even on the FA view appears to be context independent: the fitting pro-attitudes toward the object remain fitting as long as its internal features remain the same, however the external context might change (Rabinowicz and Rønnow-Rasmussen 2004, p. 409).

Rabinowicz and Rønnow-Rasmussen's suggestion that there is nevertheless a sense in which values may be non-relational according to FA analyses can be elaborated by appealing to a distinction that they make in their 1999 paper. There, they distinguish between what they call the *supervenience* base of values and the *constitutive* base of values:

The idea of a source of value is ambiguous between at least two interpretations: on the one hand, one may be thinking of the features of an object on which its final value *supervenes* (its "good-making" properties). If these features are internal to the object, i.e., non-relational, then the relevant value is intrinsic, as we are using this term. On the other hand, one may have in mind the *constitutive grounds* of an object's final value. The latter may well lie outside the object itself even though the former are internal to the object (Rabinowicz and Rønnow-Rasmussen 1999, pp. 117-118).

Thus, although FA analyses involve some relation between values and evaluators, this relation may not be the one at stake in Moore's intrinsicity principle. The former relation, it would seem, pertains to the *constitutive* base of values, while the relationality of properties at stake in Moore's principle pertains to their *supervenience* base.

This can be intuitively grasped by noting that although an FA analyst would consider that the existence of the final value of, say, a beautiful sunset partly depends upon the responses of evaluators, she would not be ready prepared to say that the sunset has its value *in virtue of* this response. The FA analyst would rather say that the beautiful sunset has its final value in virtue of the natural properties that a beautiful sunset has independently of the evaluator's evaluation. The "in virtue of" relation here points to what Rabinowicz and Rønnow-Rasmussen call the supervenience base of values. They specify that the two bases concern two distinct subfields of ethics which must not be conflated: "Claims about the supervenience bases of value belong to axiology, while claims concerning the constitutive grounds are perhaps best seen as belonging to metaethics, even though the boundary between these two disciplines is not as clear-cut as one might wish" (Rabinowicz and Rønnow-Rasmussen 1999, p. 118). Claims about the supervenience base concern axiology because that subfield develops theories about the features in virtue of which things can count as valuable, whereas claims about the constitutive base concern metaethics because that subfield constructs theories about the epistemological and ontological source of values. It is in this latter sense that FA analyses make values relational to evaluators insofar as they make values partly dependent upon evaluators' attitudinal responses. Hence, the relationality of values involved from the outset by the response-dependence of values in FA analyses does not in fact contradict Moore's intrinsicity principle. In order to assess the validity of this principle, one must examine cases where the relations upon which values depend are located in their supervenience rather than in their constitutive base. The following sections examine cases of this kind.

ARE THERE NON-INSTRUMENTAL EXTRINSIC VALUES?

Korsgaard (1983), O'Neill (1992), Kagan (1998), and Rabinowicz and Rønnow-Rasmussen (1999) discuss a series of cases where seemingly final values supervene on relational properties. A first case concerns the value that some things seem to have in virtue of their rarity. As Kagan remarks: "Many people, I think, are attracted to a view according to which the intrinsic value of an object depends in part on how rare that object is, or (in the limiting case) on its being completely unique. Obviously enough, however, uniqueness is not a property that an object has independently of whatever else may exist in the world" (Kagan 1998, p. 102). And as O'Neill notes, rarity as a ground for final value seems to have central importance for the issue of endangered species in environmental ethics:

If any property is irreducibly relational, then rarity is. The rarity of an object depends on the non-existence of other objects, and the property cannot be characterized without reference to other objects. [...] The preservation of endangered species of flora and fauna and of unusual habitats and ecological systems is a major practical environmental problem. Rarity appears to confer a special value to an object (O'Neill 1992, p. 124).

As O'Neill remarks, the value based on rarity is directly tied to another major value in environmental ethics, namely diversity, which also involves relations. O'Neill mentions yet another relational property that many environmental ethicists take as a ground for final value: "one might value wilderness in virtue of its not bearing the imprint of human activity [...] To say 'x has value because it is untouched by humans' is to say that it has value in virtue of a relation it has to humans and their activities. Wilderness has such value in virtue of our absence" (O'Neill 1992, p. 125). In a much less serious vein, Rabinowicz and Rønnow-Rasmussen (1999, p. 121) introduce the hypothetical case of the final value that a dress may have in virtue of its relational property of having belonged to Princess Diana. All these examples present cases where some objects appear to be valuable for their own sakes in virtue of relational properties. In so doing, they constitute potential counterexamples to Moore's intrinsicity principle. But are these genuine cases of relational final value?

A first indication that they are indeed is the phenomenological evidence that the evaluative attitudes experienced by the evaluators in these cases appear to be authentic attitudes of final valuation. Rabinowicz and Rønnow-Rasmussen argue on these grounds that it would be implausible to interpret such examples as cases of instrumental evaluations in disguise:

Diana's dress is perhaps valuable merely as a means: merely because it allows us to establish an indirect connection to a person we admire or find important in one way or another. Having such a connection may be something that we set a final value on. Couldn't this be what is going on here? Not necessarily. Even if the desire to establish such an "affiliation" with Diana may well be a part of the causal explanation of our

evaluative attitude towards the dress, this does not imply that the evaluative attitude itself is of the instrumental kind: if we idolise Diana, we do not simply find the dress useful for some purpose; we ascribe an independent value to it. Compare this with O'Neill's example: the wilderness is not simply instrumental in allowing us to come into contact with something (otherwise) untouched by humans. Even if we could never visit the wild area, it would still keep its value from our point of view (Rabinowicz and Rønnow-Rasmussen 1999, p. 121).

To be sure, the evaluative attitudes in these cases seem to be the very opposite of instrumentalizing ones. From the perspective of FA analyses, such phenomenological evidence should be assigned much weight, I think, given the close psychological and conceptual connection that these analyses establish between values and attitudes. Unless Moore's intrinsicity principle can be shown to have at least as much *prima facie* plausibility as the FA style of analysis, and given the striking abundance of cases similar to those presented by Korsgaard, Kagan, O'Neill and Rabinowicz and Rønnow-Rasmussen, it seems legitimate to shift the burden of proof onto the supporters of Moore's principle. The next section will show that Moore's intrinsicity principle may in fact have something closer to *prima facie* implausibility, for, as we will see, this principle requires, to earn some minimal plausibility, that its supporters accept the problematic epicycle of Moore's theory of organic unities.

CONTRIBUTIVE, SIGNATORY AND INSTRUMENTAL VALUES

Besides the phenomenological evidence yielded by paradigm examples, some additional evidence against Moore's principle may be found when considering that, as O'Neill (1992, pp. 124-125) argues, the impression that all relational values must be instrumental may just stem from an equivocal habit of language which designates two different concepts—final value and non-relational value—by the same term, “intrinsic value.” Bradley (2001, pp. 49-50; 1998, pp. 109-110) offers some resources for this purpose when he proposes that instrumental value may just be one special case among others of extrinsic value. On Bradley's proposal, besides *instrumental* value, which he defines as a value that results from a thing's “*causal* relationships with other things” (Bradley 2001, p. 49, italics mine), there exist at least two other kinds of extrinsic value: *contributive* (or *contributory*) value, i.e., “the value something has in virtue of being a part of a valuable whole,” and *signatory* value, i.e., the value something has “because of what it signifies” (Bradley 1998, pp. 110-111). On this view, the value that things have in virtue of their rarity may be labeled as *contributive* insofar as this value results from these things' contribution to the intrinsic value of a larger whole (by making it more diverse, as Bradley suggests). The value of a wilderness area's not bearing the imprint of human activity may also be viewed as *contributive* in that it results from preventing the intrinsic disvalue which would result from some human presence in it¹¹. And the value of Diana's dress may be depicted as *signatory* in that it results from what the dress represents for a Diana idolizer. Thus, on this picture, the value of rare things, of wilderness areas, and of Diana's dress would be *relational* yet without thereby being *instrumental*.

Would this confer *final* value on rare things, wilderness areas, and Diana's dress? In other words, what happens to final value according to Bradley's "de-binarization" of the relation between intrinsic and instrumental value? Is some value final once it is not instrumental, including contributive and signatory values as final ones? Or does it remain the case that only intrinsic (non-extrinsic) values can be final? This question seems undecidable on the sole ground of Bradley's taxonomy of relations. Regarding the *contributive* relation, this raises an issue brought up by Moore's well-discussed principle of organic unities. Moore developed this principle in part to explain how the same thing—e.g., beauty, which, he thought, has some final value in isolation—could generate more final value in contexts where it is contemplated and appreciated. Such an increase in final value seems to violate his intrinsicity principle, as beauty then seems capable of generating some amount of final value in virtue of a relational property, namely its property of being contemplated and appreciated. To remain consistent with his intrinsicity principle, then, Moore developed his theory of organic unities, which explains that when contemplated and appreciated, the beauty involved retains the exact same amount of final value that it has in isolation, and that the increase in final value generated by its being contemplated and appreciated belongs to the whole formed by the complex "beauty + contemplation," rather than to beauty itself. Thus, in agreement with the intrinsicity principle, the contemplation and appreciation of beauty may increase the total final value of the world, yet the final value of beauty itself remains unchanged by such a relation.

However, not everyone agrees with this interpretation of the increase in final value generated by organic unities. Hurka (1998) delineates two possible interpretations of this value, which he calls the *holistic* and *conditional* interpretations. Under the holistic interpretation, which is the one to which Moore subscribes, the final value added by the combination of parts into a whole is, as we've seen, located strictly in the whole, and so the final value of the associated parts themselves remains the same. Under the conditional interpretation, in contrast, "the intrinsic value of a state can change when it enters into a larger whole, so its value or degree of value is altered by its relations to other states" (Hurka 1998, p. 303)¹². In the case of the contemplation of beauty, this would imply that the final value of a beautiful thing is itself increased when it is contemplated. In sum, on the holistic view, there is only *upward* determination of the final value of a whole by its parts in association, while on the conditional view, there is *downward* determination of the final value of the parts by the whole formed through their association.

The question now is: Which interpretation is the correct one? In fact, both views seem plausible depending on which examples they are applied to. Bradley (2002) makes a convincing case, I think, that the final value generated by the addition of rare parts into a whole (by making it more diverse) must be interpreted holistically, that is, as belonging to the whole without affecting the parts:

Suppose, for example, that A is a beautiful painting, that B is a painting exactly like A, and that C is a beautiful piece of music. The aesthetic contemplation of A may have the same [intrinsic] value as that of B

and also the same [intrinsic] value as that of C. But the whole that is the aesthetic contemplation of A followed by that of C is intrinsically better than that whole that is the aesthetic contemplation of A followed by that of B. Hence one could say that the value of a *bonum variationis* is greater than the sum of the values of its constituent parts (Chisholm 1986, p. 71; cited in Bradley 2002, p. 34).

By Bradley's (2002, pp. 39-40) own admission, however, this accounts only for some organic unities and is therefore no proof of the *general* validity of the holistic view over against the conditionalist one. Counter to Bradley, Olson (2004) formulates a convincing defense of the conditionalist view's superiority for some other cases like that of *wicked pleasure* and *compassionate pain*. As Olson notes, the difference between the final value of "John's being pleased" and the final disvalue of "John's being pleased at Mary's pain" is better accounted for by the conditionalist interpretation, for it seems implausible to say that the final value of "John's being pleased" is the same in both cases. The total final value of "John's being pleased at Mary's pain" is lower than that of "John's being pleased" in isolation because "John's being pleased" itself is made disvaluable in a context where it is a case of wicked pleasure. And similarly for the difference between the final disvalue of "John's being pained" and the final value of "John's being pained at Mary's pain," where "John's being pained" becomes valuable (or at least less disvaluable) in a context where it is a case of compassionate pain (cf. Olson 2004, pp. 37-41). Thus the right solution to the holist/conditionalist debate about Moorean organic unities seems to be: *It depends!*

To see on what it may depend, it can be informative to look at Korsgaard's discussion of the contributive value of beauty. Criticizing Moore's own treatment of this case and defending a conditionalist interpretation of it, Korsgaard complains that

Moore's view, and the intuitionistic method of isolation, veil or obscure the internal relations within the organic unity in virtue of which the organic unity has its value. Whereas the Kantian account, which focuses on rather than ignoring the internal relations of the valuable whole, allows us to see why happiness is valuable in just this case and not in another case. Moore can only say that the combination of happiness and good will works (is a good recipe, so to speak) while happiness plus the bad will does not. Kant can say that happiness in the one case is good because the condition under which it is fully justified has been met (roughly, because its having been decently pursued makes it deserved). Those internal relations reveal the *reasons* for our views about what is valuable, while Moore's view tends to cover up these reasons (Korsgaard 1983, p. 95).

The last sentence of this passage is the most important one, for it states the rationale behind Korsgaard's preference for the conditionalist view. The conditionalist interpretation is better in this case, according to Korsgaard, because it makes the supervenience base of values more transparent—it shows the reasons

which may ground the evaluators' appropriate evaluations. This rationale resonates well with the picture given by FA analyses of the relation between values and their supervenience base. Thus it seems that Korsgaard indicates a response available to FA analysts for addressing the holist/conditionalist issue. Fitting-attitude analysts may tackle this issue in the very spirit of their approach simply by stating that whether holism or conditionalism should be the privileged interpretation should be determined on a case-by-case basis according to what our fitting evaluative attitudes respond to. Thus, FA analysts can simply state that holism should apply just when the added value grounds an evaluative response that is directed towards the whole, whereas conditionalism should apply just when the added value grounds an evaluative response that is directed towards the part. On this picture, the contributive relation between rare things and the whole to which they belong would increase the final value of the whole rather than of the part by making that whole, but not the parts, more fitting for an attitude of, say, admiration (responding to the increased variety of the whole). For instance, the contributive character of the absence of humans' imprint in a wilderness area would increase the final value of the area as a whole, but not that of the human absence itself in this context, by making the wilderness area as a whole fitting for an attitude of, say, respect. Returning to the signatory value of Diana's dress, the corresponding implication would be that the dress may acquire final value from its signatory relation insofar as this signatory relation makes it a fitting object of, say, a treasuring attitude.

Such a result, however, runs partly counter to Bradley's endeavour when he delineates kinds of non-instrumental extrinsic values. The FA analytic picture proposed here has the corollary that a thing's being in a *causal* relation with something else is not a necessary condition for this thing's having *instrumental* value. When our fitting attitudes determine the location of the final value generated in organic unities, they also thereby determine the location of some instrumental value. In the case of the value that a rare thing adds to a whole, for instance, our fitting attitudes assign some additional final value to the whole, thereby assigning *instrumental* value to the parts (regardless of the fact that this value does not supervene on causal relations). Similarly, in the case of the value that the absence of humans adds to a wilderness area, our fitting attitudes assign some final value to the area, thereby assigning *instrumental* value to the absence of humans (regardless of the fact that humans' absence *constitutes* rather than *causes* the existence of the wilderness area). This picture thus retains—at the *attitudinal level*—the dualistic opposition between final and instrumental values that Bradley sought to dissolve (though this account also acknowledges his observation that relational properties need not be causal).

But there is more besides contributive and signatory final values. Kagan and Korsgaard present a plethora of allegedly paradoxical cases where some things acquire final value in virtue of their *usefulness*, suggesting that final value may even supervene on *instrumental* relations! Kagan (1998, pp. 102-104) mentions “an elegantly designed racing car,” “excellence in various practical arts,” and “the pen used by Abraham Lincoln to sign the Emancipation Proclamation, free-

ing the slaves”; and Korsgaard (1983, p. 89) enumerates “[m]ink coats and handsome china and gorgeously enameled frying pans¹³.” In these cases as much as in those of rare things, wilderness areas and Diana’s dress, the phenomenological evidence indicates that the evaluators’ attitudes at least sometimes consist in genuine final evaluations. And the existence of such cases pleads even more strongly in favour of my proposal (*pace* Bradley) to define instrumental value not on the basis of the kinds of relations (causal, contributive, etc.) occurring in the supervenience base, but rather on the basis of the attitude befitting the thing in its relational context. In fact, the term “instrumental value” seems to have two meanings: first, a *relational* meaning (Bradley’s sense), which signifies roughly a thing’s *causal* properties, that is, its ability to cause something else; and second, an *attitudinal* meaning, which signifies a person’s fitting stance towards a thing. At the attitudinal level, final value occurs when a thing is the ultimate target of a person’s fitting evaluative attitude, and instrumental value occurs when the thing is valuable *for the sake of* something else¹⁴. With this distinction in mind, the cases of final value grounded in instrumental value can easily be explicated if one pays attention to the difference between the relations “in virtue of” and “for the sake of.” While the “in virtue of” relation, as we have seen, indicates the supervenience base of evaluative attitudes, the “for the sake of” relation indicates what one may call the *target* of evaluative attitudes. And what has been shown by cases of contributive and signatory values applies in exactly the same way to instrumental values in the relational sense: our evaluative attitudes are not dictated univocally by the kind of relations characterizing the properties *in virtue of* which they are grounded. Just as nothing precludes John’s pain from being valuable for its own sake in virtue of its contribution to an occurrence of compassionate pain, nothing precludes, say, handsome china from being valuable for its own sake partly in virtue of its usefulness. We have seen in the last paragraph that attitudinal instrumental value can supervene on types of extrinsic values that are not relationally instrumental. The independence also goes the other way around: relational instrumental value does not dictate attitudinal instrumental value.

This FA analysis-inspired treatment of contributive, signatory, and relational instrumental values thus pleads strongly in favour of the genuineness of the examples of relational final value presented by Korsgaard, O’Neill, Kagan, and Rabinowicz and Rønnow-Rasmussen. The next section examines a sophisticated attempt to defend Moore’s intrinsicity principle against these examples.

ZIMMERMAN AND OLSON’S REDUCTIONIST STRATEGY

One attempt to defend Moore’s intrinsicity principle is the reductionist strategy employed by Zimmerman (2001a; 2001b, Chap. 3) and Olson (2003). This strategy consists in arguing that the value of *objects*—such as rare things, wilderness areas and Diana’s dress—can be reduced to the value of *states of affairs* or *tropes*. On this conception, the value of Diana’s dress (an object) would be reducible to that of the state of affairs “This dress belonged to Diana” or that of the trope “This dress’s having belonged to Diana.” As states of affairs and tropes have the same implications for my purposes, I will speak uniformly of “states¹⁵.”

Zimmerman and Olson's reductionist attempt formulates a defense for Moore's intrinsicity principle by building on the observation that the *relational* properties of objects are, at the same time, *intrinsic* properties of the states in which they are involved. This observation paves the way for an indirect defense of Moore's principle: if it can be shown that what looks like the relational final value of an object can always be reduced to the state in which this object occurs, then all of the (seemingly) *relational* final values of objects will ultimately be *intrinsic* final values of states. This would confirm Moore's intrinsicity principle by making it *necessarily* true, regardless of whether the value supervenes on some properties that are relational to the objects involved in the finally valuable states.

Before examining Zimmerman and Olson's proposal, it may be relevant to consider an objection that can be addressed to it from the outset, because doing so will help us to grasp what exactly is at stake. This objection has been very well formulated by Kagan, who anticipated a reductionist attempt equivalent to the one later proposed by Zimmerman and Olson:

It might be suggested, however, that although it is a common enough practice to view objects as the bearers of intrinsic value, it is nonetheless preferable to hold that facts (or, perhaps, states of affairs) are the only genuine bearers of intrinsic value. [...] [I]t is not implausible to suggest that it is an intrinsic property of a given fact that it concerns the specific objects and properties that it does. That is, it would not be implausible to claim that it is an intrinsic property of the *fact* that there exists a pen which was used to sign the Emancipation Proclamation [...] If we accept this claim about the intrinsic properties of facts [...] and we then combine it with the earlier claim that strictly speaking only facts are the bearers of intrinsic value, then the following result emerges: one can accept the thrust of all of my examples, while still accepting the dominant philosophical tradition that intrinsic value turns solely upon intrinsic properties (Kagan 1998, p. 111).

As Kagan points out, a proposal like Zimmerman and Olson's seems to save Moore's intrinsicity principle at quite a high price: on such a proposal, Moore's principle loses its regulatory role for evaluations involving objects. So what is at stake with Zimmerman and Olson's reductionist attempt is not so much Moore's principle itself as a debate over the very possibility that objects can be genuine bearers of values. As will be seen shortly, Zimmerman and Olson's attempt leans towards the result that anytime one thinks one is valuing an object for its own sake, one is in fact valuing it somehow instrumentally. This would make many evaluations *delusively* final.

For the sake of simplicity, I will begin by focusing on Zimmerman's version of the reductionist argument and later turn to Olson's only where his version usefully supplements Zimmerman's. Zimmerman's reductionist attempt starts by making the claim that final value must be *nonderivative*:

Consider the claim that pleasure is good for its own sake. I take this to mean that every state of pleasure, every state consisting of someone's being pleased, is good simply in virtue of being such a state. There is no helpful explanation why the state is good; it just is good "as such," that is, good in virtue of its own nature. But though unhelpful, in that this account of the goodness of pleasure does not cite something else in terms of which the goodness of pleasure may be understood, it does, contrary to the first objection, provide some insight into the nature of final value. Such value is *nonderivative*; it is the ground or source of nonfinal values (such as those of charity and of hitting someone on the head with a hammer), values that may thus be declared derivative. All explanation must come to an end somewhere; the explanation of values stops with the citing of final values (Zimmerman 2001a, p. 193; see also 2001b, p. 37).

So Zimmerman's test is simple: any value that can be explained in terms of another one is *derivative*. Zimmerman's purported evidence for the claim that only those things that are nonderivatively valuable can be finally valuable is that the inexplicability of their value indicates that they are valuable "as such," or "in virtue of their own natures." Zimmerman reasons as follows: "where no helpful explanation as to why something is good is available, this is because the thing in question just is good 'as such,' that is, good in virtue of its own nature. Given that something's nature is intrinsic to it, my contention and suggestion jointly imply that *nonderivative value is intrinsic to its bearer*" (Zimmerman 2001a, p. 194; see also 2001b, p. 38). Zimmerman then mobilizes this observation against the examples presented by Korsgaard, Kagan, and Rabinowicz and Rønnow-Rasmussen:

If we were to ask these authors why the objects in question are good, we might expect, given what was said in the last paragraph, that they would simply answer, "They just are good 'as such.' They're good in virtue of their natures." But this is not what they say. Korsgaard attributes the value of the objects she mentions to their "instrumentality," that is, to their helpfulness in allowing us to accomplish certain tasks. Kagan attributes the value of Lincoln's pen to the unique historical role it played. Rabinowicz and Rønnow-Rasmussen attribute the value of Diana's dress to the fact that it belonged to Diana (Zimmerman 2001a, p. 194; see also 2001b, pp. 37-38).

I suspect that I am not the only one to remain unconvinced by Zimmerman's nonderivativity criterion. But as his whole argument hinges on it, let's grant it for now and see how Zimmerman deals with an objection raised by Rabinowicz and Rønnow-Rasmussen from the perspective of FA analyses of values:

[I]f we accept the general idea that value is what calls for an appropriate response, the response in question need not consist in just preferring or promoting. There may well be other alternatives: preference is not the only attitude to be considered, nor is promoting the only behav-

our that may be relevant in this context. [...] Alternatively, one might relate value to thing-oriented behaviours, or perhaps better, to thing-oriented attitudes-cum-behaviours: value is what we should cherish, protect or care for. [...] In fact, when one realises how many various types of responses could be relevant in this context, it becomes tempting to draw the conclusion that any monistic analysis of value in terms of one particular type of response would be inadequate. [...] Given such a pluralist approach to value analysis, the main motivation to reduce thing values to state values disappears. Valuable things may be objects that call for specific thing-oriented attitudes or behaviors: a wilderness untouched by human hands calls for protection, Diana's dress is an object to be cherished and preserved, and so on (Rabinowicz and Rønnow-Rasmussen 1999, pp. 124-125).

Zimmerman explores two possible responses to this objection¹⁶. The first consists in denying that Rabinowicz and Rønnow-Rasmussen's thing-oriented attitudes cannot also apply to states:

Notice that in many cases the attitudes that are directed toward individual objects may also be directed toward other things as well. I may love, admire, and respect someone, but I may also love, admire, and respect what he does. The fact that I have these pro-attitudes toward him doesn't itself show that he has final value, for these attitudes may derive from the attitudes I have toward what he does. I may admire what John does for its own sake (for example, I may admire his display of courage), and I may thus admire him for what he does (I may admire him for his display of courage), but this doesn't show that I admire John for his own sake (Zimmerman 2001a, p. 197; see also 2001b, p. 41).

Here, I think, Zimmerman remains unconvincing. First, it would be fairly plausible, I think, to say that talk of admiring or respecting what someone does are just elliptical ways of speaking of admiring and respecting the persons themselves (in virtue of their actions). Moreover, even assuming that these ways of speaking are literal, it remains unclear that, for instance, the admirable character of John's acts cannot—as would however seem to be the case—somehow *infuse* or *insufflate* some final value into John himself and make him admirable. In order to deal with cases where different evaluative attitudes apply to the object and the state, Zimmerman attempts his second strategy:

What exactly is it about the idea that individual objects have a different sort of value from the value that states have that requires us to say that the former value is final? Why could it not be, for instance, that, even if Diana's dress is to be treasured but its having belonged to Diana is to be valued in some other way, still the dress is to be treasured, not for its own sake, but for the sake of the state in question? That one entity derives its value from another would not appear to require that the two have exactly the same type of value (Zimmerman 2001a, p. 200; see also 2001b, p. 44).

The problematic character of this response seems even more obvious. In principle, a reduction implies that the conversion of one thing into another has no residue (no loss of information, one could say). But if Zimmerman concedes that there is some incommensurability between the input and the output of the reduction, then he seems thereby to recognize that the reduction fails due to incompleteness. Rabinowicz and Rønnow-Rasmussen make this point quite forcefully:

It seems that certain pro-attitudes that fit things or persons are not fitting or perhaps may even be impossible to hold with respect to tropes. Thus, for example, consider the attitude of respect. I can respect a person but it sounds odd, to say the least, to say that I respect a trope. I can respect Ann for her courage, but this is an attitude I hold towards Ann, and not towards her courage. [...] Conversely, some appropriate pro-attitudes towards tropes are not appropriate towards things or persons. For example, I can be exhilarated by a display of courage in Ann but it would be wrong to say that I am exhilarated by Ann, even though I may well come to value her more on account of her courage. Similarly, I can rejoice in her happiness, but I cannot rejoice in Ann herself. These examples suggest that the value of a concrete object (of a person or a thing) cannot just consist in the value of the corresponding tropes. [...] Consequently, the value of a concrete object and the value of a trope must be different from each other if the concrete object and the trope call for different responses (Rabinowicz and Rønnow-Rasmussen 2003, p. 222).

At this point, it will be illuminating to integrate some resources from Olson's version of the reductive argument. In response to Rabinowicz and Rønnow-Rasmussen, Olson (2003, p. 416) introduces a telling qualification concerning the kind of reduction pursued here. In his view, this reduction is not one of *equivalence*, but of *specification*. Thus, the reductive relation goes only one way: it says that whenever an object has value, this value is rooted in the value of a state, but it says nothing about when a state's value engenders an object's (non-final) values. This qualification may remove the requirement that the reduction have no residue. Yet the question remains: why would such rooting imply that only the state *really* has final values while the object is left with "merely *derived* values" (Olson 2003, p. 417)? Olson's answer brings us back to Zimmerman's initial contrast between derivative and nonderivative values, but through a formulation that reveals the very presupposition behind it:

Recall that we characterised final value as that which is valuable for its own sake. Now, if a person is valuable in virtue of, or because of, or for the sake of, her courage, this person is not valuable for her own sake, but rather for the sake of something else, namely her courage—the trope—which doubtless is a component of the concrete person, but a contingent one. It would be an easy thing to imagine this person lacking her courage, and thus this value (Olson 2003, p. 418).

Notice that in this passage, Olson lumps together the two types of relations that I have argued are distinct: the “for the sake of” and the “in virtue of” relations. Thus, Olson treats the relation that refers to the *target* of the evaluation as equivalent to the relation that points to its *supervenience* base. But if one does this, of course, the result will be that as soon as one values something *in virtue of* some of its features, one will thereby be valuing it *for the sake of* these features rather than *for its own sake*. This, I think, is an invalid way of denying that objects may be valued for their own sakes. For more illustration, take Olson’s ski example:

If one would go on asking the skier further questions about why he values or appreciates skiing, or what it is about skiing that he values or appreciates, one would presumably find out that there is (are) some feature(s), characteristic of skiing that the skier values or appreciates for its (their) own sake(s). The immediate object(s) of his evaluative attitude(s) would then be this (those) characteristic(s), rather than the skiing itself (Olson 2003, pp. 418-419).

Such an interpretation, I think, again conflates the *target* and the *supervenience* base of the evaluation. On Olson’s account, one really values something for its own sake only when one cannot explain *in virtue of* what features of the thing one values it. One’s being able to do so would automatically imply that one values the thing for the sake of these features rather than for its own sake. In other words, as soon as an explanation is available for a given evaluation, then the value becomes derivative and therefore nonfinal. This way of interpreting the “for the sake of” relation constitutes what Olson labels the “*strict* reading of the notion ‘value for its own sake’,” which he opposes to Rabinowicz and Rønnow-Rasmussen’s looser reading. I think that Olson’s strict reading is much too strict—indeed, *arbitrarily* strict, as will be shown shortly.

Recall that Zimmerman defends the nonderivativity criterion, which Olson here just assumes, on the grounds that applying this criterion guarantees that the value at stake really is *intrinsic to its bearer*. The criterion does this by guaranteeing that the bearer acquires its value solely in virtue of properties that are *intrinsic to its nature*. Such a defense, however, seems to conflate a *sufficient* with a *necessary* condition: nonderivativity’s guaranteeing that the value supervenes on properties intrinsic to their bearer’s nature does not imply that this happens *only when* nonderivativity occurs. In fact, the only difference between many cases of derivative value and what occurs with nonderivative value is the following: when the value is derivative, the nature of the thing in question is *transparent* to the evaluator, so that she can explain her evaluation by specifying the components of this nature that ground her evaluation; whereas when the value is nonderivative, the nature of the thing remains *opaque* to the evaluator. This opacity is, in fact, the only reason why the evaluator cannot explain her evaluation, and why her only option is then to say that she values the thing “in virtue of its own nature.” On these grounds, I take it that the derivative/nonderivative distinction does not reflect any fundamental contrast about how evaluations relate to their supervenience bases.

TOWARDS AN ESSENTIALITY PRINCIPLE?

My criticism of Zimmerman and Olson's nonderivativity criterion nonetheless concedes them something significant. It concedes them what may be called the *essentiality* principle (which must not be conflated with the *intrinsicity* principle). This principle states that a thing may only have final value in virtue of features that are *internal to its nature*. This principle underlies my rejoinder to Zimmerman and Olson that the value of a state insufflates value into an object when the state instantiates a property that is *internal to the nature* of the object. Before further explicating the essentiality principle, let's first observe how something like it seems to underlie Olson's worry that, were I to value someone for her courage, I would value her in virtue of something that is *contingent* to her (see the above quotation). Contingency seems to be the intuitive reason for Olson's worry. And there seems to be something right about this intuition: if someone values someone (or something) in virtue of properties which are contingent to that person (or thing), then it would seem right to say that the *target* of her evaluation cannot really be that person (or thing). But let's imagine that, in Olson's example, the person's display of courage can be interpreted as an enduring character trait, as a reflection of her deep personality, or of the internal nature of the person she is, so to speak. In such a case, then, it seems that Olson's worry disappears, and consequently that it is perfectly unproblematic to say that the person who is valued in virtue of her courage is valued for her own sake. And this is exactly what the essentiality principle predicts: insofar as this person is valued in virtue of features belonging to her own nature, she can be valued finally. Now take Olson's skiing example. I doubt that the features in virtue of which skiers enjoy skiing are so contingent to this activity; if there were no stable features of skiing that skiers are confident to find again and again when they go hurtling down the slopes, their general attitude towards skiing would be very hard to explain. If these stable features can be seen as part of the internal nature of skiing, however, it seems perfectly correct to say that the skiers value skiing for its own sake. Contrast this with a case where some people think that they enjoy skiing for its own sake in virtue of its being a good occasion to display their economic wealth by wearing ostentatiously expensive equipment. Presumably, such a display is not part of the internal nature of skiing, and so the essentiality principle would entail that these show-offs' evaluation is *delusively* final. And consider the case where I admire a butterfly so much in virtue of its beauty that I can't resist suffocating it so that I can preserve it forever in my collection. In such a case, my evaluative attitude is conceivably one of final evaluation: I value the butterfly *as an end* in virtue of its beauty. But given that I am sacrificing its life, presumably a property that is dearer to a butterfly's nature than is its beauty, the essentiality principle states that my attitude, like that of the show-off skiers, would have to be qualified as *delusively* final. To me, these upshots of the essentiality principle seem fairly intuitive. Although FA analysts want to give much credit to attitudinal responses, they should nonetheless, it seems to me, leave some room for a notion of *delusively final* evaluative attitudes (which are in fact concealed acts of instrumentalization).

Although, as I said, the essentiality principle is not equivalent to Moore's intrinsicity principle, it nevertheless has some kinship with it (see Bradley 2002, pp. 23-24; Rabinowicz and Rønnow-Rasmussen 1999, p. 116, p. 127). First, it seems to be a principle that Moore himself endorsed at times:

When I say, with regard to any particular kind of value, that the question whether and in what degree anything possesses it depends solely on the *intrinsic nature* of the thing in question, I mean to say [...] that it is impossible for what is strictly one and the same thing to possess that kind of value at one time, or in one set of circumstances, and not to possess it at another; and equally possible for it to possess it in one degree at one time, or in one set of circumstances, and to possess it in a different degree at another, or in a different set (Moore 2000, pp. 260-261, italics mine).

Moreover, it seems thinkable to interpret Moore's isolation test as aimed primarily at isolating *essential* properties. Given that states of affairs have their intrinsic properties necessarily, it follows that Moore's isolation test, when isolating the intrinsic properties of *states*, thereby isolates their essential ones. This equivalence does not, however, apply in the case of objects, for, as Rabinowicz and Rønnow-Rasmussen note, "[a] concrete individual, such as, say, Theseus' ship, may well have internal properties that are contingent, say, such features as its colour, or the property of containing as a part a particular plank, *a*" (Rabinowicz and Rønnow-Rasmussen 1999, p. 116). But Moore's intrinsicity principle may have in fact been meant to serve as an essentiality principle, intended to ensure that objects' value supervene not on their intrinsic but on their essential properties. Indeed, a remark by Rabinowicz and Rønnow-Rasmussen suggests that this interpretation might make Moore's views more intuitive on this matter:

[T]he intrinsic value of an object may well vary in different possible worlds, if this value supervenes on the object's *contingent* internal properties. In other words, one should not confuse intrinsic value with *essential value*, where the latter is the value that the object has by necessity, i.e., in every possible world in which it exists. This suggests, by the way, that the notion of an intrinsic value may not be as normatively interesting as many have thought. After all, what is so special about value that supervenes on the object's internal rather than relational properties, especially if the former may be just as contingent to the object as the latter? One can easily see the normative relevance of the notion of a final value (at least if such value is analysed in terms of a range of the fitting responses that the object calls for), but the concept of an intrinsic value seems to lack a special normative interest (Rabinowicz and Rønnow-Rasmussen 1999, p. 127).

At the very least, my discussion of Zimmerman and Olson's reductionist attempts suggests that an identification of final value with essential value has more *prima facie* plausibility than one with intrinsic value.

It must be noted that the essentiality principle is not necessarily more demanding than the intrinsicity principle. As far as objects are concerned, essential properties can very often be relational. For instance, many objects to which functions are assigned, like organs and tools, are defined by their causal relations with their bearers or users. Arguably, a spontaneously generated hammer on an uninhabited exo-planet is not really a hammer (see McLaughlin 2002, pp. 135-136), and likewise for a spontaneously generated orphan heart. Similarly, many symbolic objects, such as dollar bills and photographs, are defined by their signatory relations. And likewise, many things that exist as parts of wholes, like book chapters and train stations, are defined by their contributory relations. All of these cases involve relational essential properties, which, should they be taken as supervenience bases for some evaluations, would, under the essentiality principle, ground genuine final evaluations. Returning to the examples supplied by Kagan, O'Neill, and Rabinowicz and Rønnow-Rasmussen, adding a rare part to a diverse whole can make it more finally admirable *qua* diverse whole, the absence of humans in a wilderness area can make the area deserve more respect for its own sake *qua* wilderness, and a dress's having belonged to Diana can make it more finally treasurable *qua* something having belonged to a celebrity.

Admittedly, this way of applying the essentiality principle is quite liberal, as the “*qua*-clause” is malleable enough to turn any instrumental evaluation into a final one just by redefining a given token as belonging to a different kind. As liberal as it is, I think that this way of applying the principle retains a regulative role with crucial importance for delineating genuine and delusive final evaluations. As I just said, one may well treasure Diana's dress, in virtue of its signatory relation with the princess, for its own sake *qua* something that has belonged to a celebrity. But according to the essentiality principle as I apply it, one may not, in virtue of this relation, treasure the dress *qua* dress. Thus in this case, *qua* dress, the dress is instrumentalized for the sake of being an object that has belonged to a celebrity. Similarly, one may well, in virtue of its not bearing any human imprint, respect a wilderness area for its own sake *qua* instantiation of pure otherness, say. But according to the essentiality principle, one may not, in virtue of this feature, value it *qua* healthy ecosystem, which is not conceptually equivalent to wilderness. If however, the wilderness of the ecosystem is taken as signatory of its health, then the ecosystem can be valued for its own sake *qua* healthy ecosystem, in virtue of its condition of wilderness. But this applies to cases of values grounded in intrinsic properties as well. Suppose that I like my neighbourhood, “La Petite Patrie,” in virtue of the fact that most of my friends live there. Although in this case my friends' presence is an intrinsic property of the neighbourhood (my friends are constituents of it along with its streets, schools, and so on), it would nevertheless be incorrect to say that I like “La Petite Patrie” for its own sake in virtue of the fact that my friends live there. What I like for its own sake, in this case, is the neighbourhood where my friends live. And let me introduce one last case, one in which a person's evaluation undermines the nature of the thing valued. Recall my butterfly example: I may well admire a butterfly for its own sake *qua* collectible object solely in virtue of its beauty, a property which presumably belongs to its nature, yet taking such a restricted part of its nature as

the supervenience base of my evaluation somehow undermines what the butterfly is. This leads me to adopt an attitude (and a consequent behaviour) which is not responsive to the full nature of the butterfly *qua* living being, and so involves some degree of instrumentalization. The regulative role of the essentiality principle, in short, is to uncover the cases of delusively final evaluations, or to put it more sombrely, to unmask cases of instrumentalization in disguise.

CONCLUSION

This paper has examined the plausibility of G. E. Moore's intrinsicity principle from the perspective of fitting-attitudes analyses (FA analyses) of values. First, by recalling Rabinowicz and Rønnow-Rasmussen's distinction between the supervenience and constitutive bases of values, it argued that although FA analyses portray values as being partly constituted by the responses of evaluators, this relation does not make the ensuing value relational in a way that contradicts Moore's principle. Secondly, the paper introduced Korsgaard, Kagan, O'Neill, and Rabinowicz and Rønnow-Rasmussen's paradigm examples of objects which intuitively have relational final value, and argued that FA analysts should be responsive to the phenomenological evidence that these cases are indeed genuine. Then, turning to Bradley's distinction between instrumental, contributive, and signatory value, the paper argued that the relevant contrast between final and instrumental values must be situated at the attitudinal level rather than at the level of the relations occurring in the supervenience base, and furthermore that this allows for genuine cases of final value in virtue of contributive, signatory, and instrumental relations. Fourthly, the paper examined Zimmerman and Olson's reductionist argument, which contends that only states and not objects can be proper bearers of final value, and proceeded to show that these authors' nonderivativity criterion is arbitrary. Lastly, the paper explored the *prima facie* plausibility of the 'essentiality principle' as an alternative to Moore's intrinsicity principle. The latter exploration, indeed, offered just a first glance at the essentiality principle along with a cursory examination of its plausibility. Many issues still have to be tackled, but I hope that this paper has indicated some prospects for fruitful further investigations.

NOTES

- * The author would like to thank Christine Tappolet, the audience at the “Colloque Jeunes Chercheurs” of the SoPhA and at the SPQ 2014 symposium “La normativité: découverte ou invention,” and two anonymous referees for helpful comments. He would also like to thank the *Groupe de recherche sur la normativité* (GRIN), the *Centre interuniversitaire de recherche sur la science et la technologie* (CIRST) and the *Fonds de recherche – Société et Culture* (FRQ-SC) for doctoral grants.
- ¹ I take the phrase “*formal* axiology” from Olson (2003, p. 413). The term can be seen as denoting the sub-field of axiology, which focuses on the nature of final value—what final value *is*—in contrast to *substantial* axiology, which attempts to determine what final values *are*.
 - ² However, Rabinowicz and Rønnow-Rasmussen (1999, p. 116) note that, rigorously speaking, something may have “internally relational” properties, that is, “properties that it possesses in virtue of its relations to its own *parts*.” As it seems clear that by “relational properties” Moore meant “externally relational properties,” I follow Rabinowicz and Rønnow-Rasmussen’s use and employ the former term in a restricted sense that includes only internally relational properties.
 - ³ The second “error” alluded to by Moore concerns his “principle of organic unities,” to which I will return below.
 - ⁴ In the following, I use the word “thing” in a very broad sense that may include all kinds of ontological entities (like “states of affairs,” “objects,” “tropes,” etc.) when I wish to remain neutral regarding the debate over the proper bearers of value.
 - ⁵ Many of the publications involved in this debate have been usefully anthologized in Rønnow-Rasmussen and Zimmerman (2005).
 - ⁶ For more detailed introductions to FA analyses, see Rabinowicz (2013), Jacobson (2011) and Rabinowicz and Rønnow-Rasmussen (2004).
 - ⁷ Although most FA analysts interpret the fittingness relation between the attitude and its object in *deontic* terms, that is, as a relation that states the type of attitude that evaluators *ought* to adopt or *have reasons* to adopt, there are also some FA analysts who interpret this relationship as a *representational* one, that is, a relation that states the attitude that evaluators are *correct* to adopt, that represents the objects at stake *as they are* (cf. Tappolet 2011).
 - ⁸ Although some find anticipations in the earlier work of Scottish moral sense theorists like Hume and Hutcheson, the historical origin of FA analyses is usually traced back to Brentano (1902) and Ewing (1947). The theory has recently experienced a revival ever since Scanlon (1998) has turned the metaethical community back to it. See Rabinowicz and Rønnow-Rasmussen (2004, pp. 394-400) for a historical survey and Tappolet (2011, n1) for a bibliography.
 - ⁹ There is of course a challenge here, and much of the disagreement among FA analysts has been over which formulation of the fittingness relation can best meet this challenge (as an anonymous referee has remarked, few proponents of FA analyses have in fact offered such naturalistic analyses).
 - ¹⁰ Of course, objections invoking the “wrong kind of reasons” problem have been a recurring worry for FA analyses (cf. Stratton-Lake 2005; Rabinowicz and Rønnow-Rasmussen 2004; Crisp 2000; D’Arms and Jacobson 2000), but some promising solutions to this problem have recently been put forward (e.g. Danielsson and Olson 2007).
 - ¹¹ See Bradley’s (2002, pp. 30-33) defense of the legitimacy of such seemingly strange counterfactual application of the notion of contributive value.
 - ¹² Hurka speaks of “intrinsic” rather than “final” value, but the context of his discussion, which leaves open to discussion the question whether “intrinsic” value can supervene on relational properties, makes clear that the concept he has in mind is the one I call “final value.” Kagan (1998) and most environmental philosophers follow this (often confusing) terminological practice. Also, Hurka speaks of *states*, but his two interpretations, as we will see, can also be applied to other kinds of ontological entities like objects.

- ¹³ I would however be more inclined to say that the final value of the Abraham Lincoln pen is explained by its signatory relation with the Emancipation Proclamation rather than by its usefulness.
- ¹⁴ This distinction is very close to the one formulated by Rønnow-Rasmussen (2002), however, I think that my formulation in terms of relational vs. attitudinal instrumental values better emphasizes the locus of the contrast than his in terms of strong vs. weak instrumental values.
- ¹⁵ The main difference is that states of affairs are *abstract predications* of properties, e.g. “a having p,” whereas tropes are *concrete* (spatiotemporally located) *instantiations* of properties, e.g. “a’s having p,” (Rabinowicz and Rønnow-Rasmussen 2003, pp. 218-219).
- ¹⁶ As Zimmerman accepts the FA analysis of value, simply ignoring Rabinowicz and Rønnow-Rasmussen’s objection is not an option available to him (see Zimmerman 2001a, p. 197; 2001b, p. 40).

REFERENCES

- Bradley, Ben, "Extrinsic Value," *Philosophical Studies*, vol. 91, no. 2, 1998, pp. 109-126.
- Bradley, Ben, "The Value of Endangered Species," *Journal of Value Inquiry*, vol. 35, no. 1, 2001, pp. 43-58.
- Bradley, Ben, "Is Intrinsic Value Conditional?," *Philosophical Studies*, vol. 107, no. 1, 2002, pp. 23-44.
- Brentano, Franz, *The Origin of Our Knowledge of Right and Wrong*, Westminster, A. Constable, 1902.
- Chisholm, Roderick M., *Brentano and intrinsic value*, Cambridge/New York, Cambridge University Press, 1986.
- Crisp, Roger, "Value ... And What Follows by Joel Kupperman New York: Oxford University Press," *Philosophy*, vol. 75, no. 3, 2000, pp. 452-462.
- D'Arms, Justin, and Daniel Jacobson, "Sentiment and Value," *Ethics*, vol. 110, no. 4, 2000, pp. 722-748.
- Danielsson, Sven, and Jonas Olson, "Brentano and the Buck-Passers," *Mind*, vol. 116, no. 463, 2007, pp. 511-522.
- Ewing, Alfred C., *The definition of good*, New York, NY, Macmillan co., 1947.
- Hurka, Thomas, "Two Kinds of Organic Unity," *Journal of Ethics*, vol. 2, no. 4, 1998, pp. 299-320.
- Jacobson, Daniel. "Fitting Attitude Theories of Value," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, 2011, URL: <http://plato.stanford.edu/archives/spr2011/entries/Fitting-attitude-theories/>.
- Kagan, Shelly, "Rethinking Intrinsic Value," *Journal of Ethics*, vol. 2, no. 4, 1998, pp. 277-297; cited as reprinted in Rønnow-Rasmussen, Toni, and Michael J. Zimmerman (eds.), *Recent Work on Intrinsic Value*, Dordrecht, Springer, 2005, chap. 9.
- Korsgaard, Christine M., "Two Distinctions in Goodness," *Philosophical Review*, vol. 92, no. 2, 1983, pp. 169-195; cited as reprinted in Rønnow-Rasmussen and Zimmerman (eds.), *Recent Work on Intrinsic Value*, Dordrecht, Springer, 2005, chap. 8.
- Mackie, John L., *Ethics: Inventing Right and Wrong*, Harmondsworth, NY, Penguin Books, 1977.
- McLaughlin, Peter, "On Having a Function and Having a Good," *Analyse and Kritik*, vol. 24, no. 1, 2002, pp. 130-143.
- Moore, George E., *Principia Ethica*. Amherst, N.Y., Prometheus Books, 1988.
- Moore, George E., *Philosophical Studies*, London, Routledge and Kagan Paul, 2000.
- O'Neill, John, "The Varieties of Intrinsic Value," *The Monist*, vol. 75, no. 2, 1992, pp. 119-137.

Olson, Jonas, "Revisiting the Tropic of Value: Reply to Rabinowicz and Rønnow-Rasmussen," *Philosophy and Phenomenological Research*, vol. 67, no. 2, 2003, pp. 412-422.

Olson, Jonas, "Intrinsicism and Conditionalism About Final Value," *Ethical Theory and Moral Practice*, vol. 7, no. 1, 2004, pp. 31-52.

Rabinowicz, Wlodek, "Value, Fitting attitude Account Of," in *International Encyclopedia of Ethics*. Blackwell Publishing Ltd, 2013, URL: <http://onlinelibrary.wiley.com/doi/10.1002/9781444367072.wbiee311/abstract>.

Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen, "A Distinction in Value: Intrinsic and for Its Own Sake," *Proceedings of the Aristotelian Society*, vol. 100, no. 1, 1999, pp. 33-51; cited as reprinted in Rønnow-Rasmussen and Zimmerman (eds.), *Recent Work on Intrinsic Value*, Dordrecht, Springer, 2005, chap. 10.

Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen, "Tropic of Value," *Philosophy and Phenomenological Research*, vol. 66, no. 2, 2003, pp. 389-403; cited as reprinted in Rønnow-Rasmussen and Zimmerman, *Recent Work on Intrinsic Value*, Dordrecht, Springer, 2005, chap. 17.

Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen, "The Strike of the Demon: On Fitting Pro-attitudes and Value," *Ethics*, vol. 114, no. 3, 2004, pp. 391-423.

Rønnow-Rasmussen, Toni, "Instrumental Values - Strong and Weak," *Ethical Theory and Moral Practice*, vol. 5, no. 1, 2002, pp. 23-43.

Rønnow-Rasmussen, Toni, and Michael J. Zimmerman (eds.), *Recent Work on Intrinsic Value*, Dordrecht, Springer, 2005.

Scanlon, Thomas M., *What We Owe to Each Other*, Cambridge, MA, Belknap Press of Harvard University Press, 1998.

Stratton-Lake, Philip, "How to Deal with Evil Demons: Comment on Rabinowicz and Rønnow-Rasmussen," *Ethics*, vol. 115, no. 4, 2005, pp. 788-798.

Tappolet, Christine, "Values and Emotions: Neo-Sentimentalism's Prospects," in Carla Bagnoli (ed.), *Morality and the Emotions*, Oxford, Oxford University Press, 2011.

Zimmerman, Michael J., "Intrinsic Value and Individual Worth," in Dan Egonsson, Jonas Josefsson, Björn Petersson, and Toni Rønnow-Rasmussen (eds.), *Exploring practical philosophy: from action to values*, Aldershot, Ashgate, 2001a, pp. 123-138; cited as reprinted in Rønnow-Rasmussen, Toni, and Michael J. Zimmerman (eds.), *Recent Work on Intrinsic Value*, Dordrecht, Springer, 2005, chap. 15.

Zimmerman, Michael J., *The Nature of Intrinsic Value*, Lanham, MD., Rowman & Littlefield, 2001b.

ESSENCE ET FONDATION

PABLO CARNINO

DOCTORANT EN PHILOSOPHIE, UNIVERSITÉ DE GENÈVE

RÉSUMÉ :

L'orthodoxie dans la littérature florissante au sujet de la fondation (*grounding*) suggère que cette notion ne peut être analysée ou exprimée en terme d'aucune autre. Par ailleurs, le primitivisme à propos de l'essence est considéré comme très plausible depuis l'article influent de Kit Fine à ce sujet¹. Cela contraint les philosophes qui emploient ces deux notions à accepter une position doublement primitiviste. Mon objectif principal est de proposer une définition de la fondation en terme d'essence. Je commencerai par présenter la conception positive en question, puis je montrerai la façon dont elle répond à un certain nombre d'objections à son encontre. L'apport d'une telle approche, si elle est couronnée de succès, devrait être évident : les philosophes qui acceptent la fondation comme notion primitive n'auront plus à le faire, pour autant qu'ils emploient déjà la notion d'essence. Pour ceux qui refusaient de parler de fondation sous prétexte que la notion était indéfinissable, la présence d'une définition éliminera *ipso facto* leur raison d'être sceptique.

ABSTRACT:

Orthodoxy in the flourishing literature about grounding has it that the notion cannot be analysed with or expressed in terms of any other notion. Similarly, a form of primitivism is taken to be quite plausible about the notion of essence ever since Kit Fine's seminal article on the topic. This forces philosophers (Kit Fine, *inter alia*) that use the two notions into a doubly primitivist position. My main aim here is to put forward a definition of grounding in terms of essence. I will first present the positive account and then show how it can handle a variety of objections. The upshot of such an approach if successful should be obvious: the philosophers that accept grounding as a primitive notion in their theorizing need not do so anymore, at least provided that they also helped themselves to a notion of essence (or objectual definition). For those who were reluctant to use grounding talk because it was allegedly undefinable my defining grounding in terms of essence eliminates *ipso facto* their reason for being skeptic.

INTRODUCTION

Une célèbre notion en métaphysique indique ce qui fait d'une chose ce qu'elle est. Parler des aspects qui font d'une chose *ce qu'elle est*, ou qui déterminent son identité, revient à parler de l'*essence* de la chose en question. Une notion moins célèbre, mais récemment très exploitée en métaphysique, transmet l'idée selon laquelle certains faits obtiennent en vertu d'autres faits. On parle de *fondation*² quand on cherche à savoir quels faits obtiennent *en vertu de*³ quels autres faits. Pour ne donner que quelques exemples : de nombreux philosophes sont d'avis que si une action est bonne, elle l'est en vertu du fait qu'elle maximise l'utilité, le bonheur ou la satisfaction des préférences. On dit aussi que si le vase est fragile, il l'est en vertu de sa composition moléculaire, ou même de sa structure atomique. Cette relation explicative spécifique est adoptée comme primitive par la plupart des philosophes qui théorisent à son sujet⁴. L'orthodoxie veut que les candidats naturels pour définir la fondation, tels que la survenance, les contre-factuels ou encore l'essence, soient voués à l'échec. Je remets en question ce dictat en présentant une définition de la fondation en terme d'essence. La définition que j'avance est tirée d'un article de Fabrice Correia⁵ dans lequel il présente une série de définitions possibles de la fondation en terme d'essence et les évalue en considérant certaines des objections que je présenterai. Je précise toutefois que la définition en question n'est pas officiellement acceptée par Correia. Kit Fine⁶ a suggéré qu'une telle définition était problématique pour plusieurs raisons. Je montrerai comment la définition proposée résiste aux objections avancées par Fine. Ensuite, j'explorerai de nouvelles objections à l'encontre de la définition proposée. Je conclurai que cette définition est en tout cas une excellente candidate et que son plus sérieux défaut semble être qu'elle implique le nécessitarisme de la fondation, une thèse qui, bien qu'extrêmement répandue chez les théoriciens de cette notion, souffre de quelques contre-exemples dont je discuterai le bien-fondé.

L'apport d'une telle définition, si celle-ci est correcte, devrait être évident : les philosophes qui acceptent la fondation comme notion primitive n'auront plus à le faire, pour autant qu'ils emploient déjà la notion d'essence. Pour ceux qui refusaient de parler de fondation sous prétexte que cette notion était indéfinissable, la présence d'une définition éliminera *ipso facto* leur raison d'être sceptique.

Avant d'aborder la définition qui m'intéressera tout au long de cet article, certaines clarifications s'imposent concernant les rapports entre essence et fondation. Ces notions ont en tout cas un point commun évident qui peut laisser entendre que définir l'une en terme de l'autre est une bonne stratégie; je veux bien sûr parler de leur force modale. En effet, si un objet a une propriété essentielle, il est alors métaphysiquement nécessaire que cet objet ait cette propriété. De plus, si x fonde y , alors l'implication « si x alors y » est métaphysiquement nécessaire. Ainsi ces deux notions entretiennent un lien systématique avec la notion de nécessité métaphysique. Par ailleurs, plusieurs théoriciens de la fondation ont exprimé, d'une manière ou d'une autre, l'idée selon laquelle la fon-

dation était liée à l'essence⁷. Ces philosophes explorent de manières diverses les rapports entre une notion explicative particulière et une notion définitionnelle spécifique. D'une manière générale, il me semble que ces philosophes sont tentés par l'idée voulant que les chaînes d'explications doivent s'arrêter quelque part ou être soutenues par quelque chose, et que la nature des choses, c'est-à-dire leur essence, est une excellente candidate pour jouer ce rôle. Une autre raison de penser qu'une approche de la fondation qui exploite la notion d'essence est une bonne piste est que plusieurs des exemples de fondation semblent évoquer des connexions dans la nature de certaines propriétés. Par exemple, il y a un lien entre la nature de la propriété d'être coloré et celle d'être rouge : le premier est un déterminable et le second un déterminé de ce même déterminable. Ce lien est d'une certaine manière la base de la relation de fondation entre le fait que la balle est colorée et le fait qu'elle est rouge. Évidemment, constater ce genre de lien n'est pas une raison suffisante pour adopter la définition que je vais proposer, puisque Fine rejette cette définition tout en acceptant le lien systématique entre essence et fondation.

DÉFINITION ET PRÉCISIONS

La définition qui va suivre exprime la fondation à l'aide d'un prédicat relationnel flanqué d'expressions pour des faits⁸, comme suit : « le fait que Sam est triste est fondé dans le fait que le cerveau de Sam est dans tel ou tel état. » La thèse selon laquelle cet usage est le mode fondamental d'expression de la fondation est appelée *prédicationisme* et s'oppose à l'*opérationisme*⁹ qui exploite, comme son nom l'indique, un opérateur flanqué de phrases, comme suit : « Sam est triste parce que son cerveau Sam est dans tel et tel état. » Mon usage du prédicat à la place de l'opérateur ne représente aucun engagement officiel, et j'espère pouvoir rester neutre sur la question de la forme canonique de la fondation¹⁰. La grande majorité des choses que je dirai s'applique également, moyennant quelques modifications, à une définition opérationiste qui aurait une structure similaire :

D_{pred} f est fondé dans f_i, f_{ii}, \dots

ssi (déf.)

f_i obtient & f_{ii} obtient & \dots ,¹¹ et

ii) L'implication suivante fait partie de la nature de f : si f_i obtient & f_{ii} obtient & \dots , alors f obtient.

La définition utilise f, f_i, f_{ii}, \dots comme des noms arbitraires pour des faits. Mais il est assez simple de voir comment (D_{pred}) s'applique à des cas spécifiques. Prenons le rapport entre un fait moral et les faits non moraux correspondants. Le fait que l'action de Sam est bonne est fondé, diront certains, dans le fait qu'elle maximise l'utilité. Ce lien de fondation signifie, par (D_{pred}), (a) que le fait que l'action de Sam maximise l'utilité obtient et (b) que l'implication suivante fait partie de la nature du fait que l'action de Sam est bonne : si le fait que l'action de Sam maximise l'utilité obtient, alors le fait que l'action de Sam est bonne

obtient. La première condition semble tout à fait innocente, pour autant qu'on ne conçoive pas l'obtention d'un fait comme une redondance nocive. Et, le cas échéant, on peut aisément remplacer les faits par des états de choses¹² pour éviter la redondance.

La deuxième clause, quant à elle, requiert l'application de la notion d'essence à des faits. En principe, une fois que l'on accepte l'idée que les objets ont une essence et que les faits sont des objets, l'idée que les faits ont une essence s'ensuit¹³. Mais il y a quand même quelques considérations en faveur de l'idée selon laquelle même des objets spéciaux comme les faits ont une essence. Il y a en tout cas deux raisons d'accepter ce type d'essence. La première est que, comme je l'ai dit, citer l'essence de Socrate, c'est mentionner les aspects de Socrate qui font de lui ce qu'il est. Ainsi, si je peux distinguer un fait, par exemple f_{iv} , d'un autre fait, par exemple f_v , alors il semble qu'il y ait des aspects du fait f_{iv} qui font de lui ce qu'il est. Une autre raison est simplement notre intuition selon laquelle les faits, à l'instar des objets, ont des propriétés plus importantes que d'autres, dans un certain sens. On dit par exemple que même si Socrate a la propriété d'être humain et celle d'être philosophe, l'une de ces propriétés est attachée à lui « plus intimement »¹⁴. De même, f_{iv} possède à la fois la propriété d'obtenir si la proposition correspondante est vraie, et celle d'être évoqué dans ce texte, mais la première propriété lui semble « plus intimement » attachée. Pour prendre un exemple plus concret : le fait que Barack Obama est le président des États-Unis a la propriété, entre autres, de me faire plaisir (admettons) et il a également la propriété de parler de Barack Obama. Mais la première propriété ne semble pas aussi intimement liée au fait en question que la deuxième. En résumé, le genre d'intuitions qui nous motivent à parler de la nature des objets se retrouve à propos de la nature des faits. Par ailleurs, Fine accepte explicitement la notion d'essence d'un fait dans l'extrait¹⁵, alors qu'il argumente contre une définition du genre de (D_{pred}) .

LES OBJECTIONS DE FINE

Fine oppose quatre objections principales au projet de définir la fondation en terme d'essence. La première est simplement qu'une telle définition ne saurait distinguer le fondement¹⁶ pluriel f_{iv}, f_v du fondement conjonctif singulier $(f_{iv} \& f_v)$. Un tel manque de capacité expressive serait bien sûr problématique, car on pense typiquement que $(f_{iv} \& f_v)$ est fondé dans f_{iv}, f_v . Par substitution des fondements pluriels par le fondement conjonctif, on obtiendrait $(f_{iv} \& f_v)$ est fondé dans $(f_{iv} \& f_v)$, ce qui constituerait une violation du principe d'irréflexivité¹⁷. Mais ce problème ne se présente simplement pas pour (D_{pred}) , car l'utilisation du terme de succès « obtient » permet de distinguer le fondement singulier « $(f_{iv} \& f_v)$ obtient » des fondements pluriels « f_{iv} obtient » et « f_v obtient ».

Par ailleurs, pour que le résultat indésirable « $(f_{iv} \& f_v)$ est fondé dans $(f_{iv} \& f_v)$ » soit dérivé de (D_{pred}) il faudrait, par application de la deuxième clause, que l'implication suivante fasse partie de la nature du fait que $(f_{iv} \& f_v)$: si $(f_{iv} \& f_v)$ obtient, alors $(f_{iv} \& f_v)$ obtient. Il faut noter que le problème se généraliserait alors à tous les faits qui obtiennent, et que (D_{pred}) aurait pour conséquence désas-

treuse la réflexivité de la fondation, ce qui impliquerait une foule de contre-exemples à l'asymétrie. Mais ces considérations sont étroitement liées à la deuxième objection, et puisque la réponse sera la même, j'introduis la deuxième objection sans attendre : une définition de la fondation en terme d'essence prédira que $(f_{iv} \& f_{iv})$ fonde f_{iv} , parce que l'implication suivante fait partie de la nature de f_{iv} : si $(f_{iv} \& f_{iv})$ obtient, alors f_{iv} obtient. Une manière naturelle et peut-être suffisante de répondre à ces objections est que l'on voit mal pourquoi la nature d'un fait atomique comme f_{iv} contiendrait des informations sur des faits conjonctifs. Parallèlement, on voit mal en quoi il relève de la définition d'un fait conjonctif comme $(f_{iv} \& f_v)$ que s'il obtient, alors il obtient. Si l'on accepte ce genre d'essentialité, pourquoi ne pas dire que c'est dans la nature de Socrate que s'il existe, alors il existe, ou encore que s'il existe, alors il existe *et* il existe? Ma première réaction est la suivante : si ces attributions concernant la nature d'un individu sont inacceptables, elles le sont tout autant à propos de faits individuels. L'implication « si Sam est triste alors Sam est triste » a bien plus à voir avec le fonctionnement de « si... alors... » qu'elle n'a à voir avec le fait que Sam est triste; ce fait est évidemment à propos de Sam et de sa tristesse. En laissant de côté les intuitions un instant, l'objection de Fine peut simplement être comprise comme si elle présumait que la notion d'essence en jeu dans (D_{pred}) inclut l'essence dérivative. On distingue habituellement l'essence constitutive¹⁸ de l'essence dérivative : alors que la première définit l'objet directement pour ce qu'il est, la seconde est attribuée à l'objet par « dérivation » sur la base de la première :

The constitutive essence is directly definitive of the object, but the consequential essence is only definitive through its connection with other properties¹⁹.

Sur la base de cette distinction, les deux prémisses de l'objection de Fine peuvent être formulées.

- (a) La notion d'essence utilisée dans (D_{pred}) est dérivative.
- (b) Si Δ fait partie de l'essence de x et que $\Delta \rightarrow \Gamma$, alors Γ fait partie de l'essence de x (c'est-à-dire que la notion d'essence dérivative est close sous la conséquence logique).

Il y a une manière assez simple de résister aux objections en question : on peut stipuler que la notion d'essence employée dans (D_{pred}) est celle d'essence constitutive. Cela bloque les objections, dans la mesure où des implications telles que « si $(f_{iv} \& f_{iv})$ obtient, alors f_{iv} obtient » et « si $(f_{iv} \& f_v)$ obtient, alors $(f_{iv} \& f_v)$ obtient » ne sont pas dans l'essence constitutive de choses comme f_{iv} et $(f_{iv} \& f_v)$. Mais pourquoi exclurait-on de telles implications de l'essence constitutive des faits en question? Supposons que f_{iv} est le fait que Sam est triste. Des implications triviales du genre : si f_{iv} obtient alors $(f_{iv} \& f_{iv})$ obtient, ou encore, si f_{iv} obtient alors $(f_{iv} \& f_v)$ obtient ne font intuitivement pas partie de ce qui définit directement f_{iv} . Cela parce que ce genre d'implications a davantage à

voir avec, par exemple, les propriétés du concept de conjonction ou encore d'implication qu'avec un quelconque fait particulier à propos de Sam. Une lecture pour le moins disponible au finéens est de dire que ces implications font partie de l'essence dérivée de f_{iv} et du concept d'implication, ou de f_{iv} et du concept de disjonction *pris ensemble*. Évidemment, je ne prétends pas que ces considérations résultent logiquement de la distinction finéenne entre essence constitutive et essence dérivative. En revanche, elles en sont au moins une interprétation plausible qui a l'avantage d'éviter l'objection que l'on considère ici.

La troisième objection concerne la distinction entre existence *simpliciter* et existence à un temps. D'après Fine, la conjonction de (D_{pred}) et de la proposition suivante entraîne une violation de l'asymétrie de la fondation :

[...] it is essential to any object that exists in time that it exists simpliciter iff it exists at a time. But the definition will then give us an equal right to say that the object exists simpliciter in virtue of existing at a time and that it exists at a time in virtue of existing simpliciter²⁰.

Toutefois, on remarque que Fine parle ici de l'essence d'objets qui existent dans le temps. Or (D_{pred}) quantifie sur l'essence de faits, pas sur la nature des objets contenus dans ces faits. Autrement dit, puisque les prémisses de l'argument portent sur un objet uniquement, et la conclusion sur le rapport entre certains faits et d'autres faits, l'argument ne semble pas valide. Afin que l'absence de lien logique soit plus claire, voici les deux propositions dont Fine a besoin comme prémisses pour conclure qu'il y a effectivement symétrie d'une relation de fondation entre le fait qu'un objet existe à un temps et le fait qu'il existe *simpliciter* :

P_1 . Cela fait partie de la nature du fait que x existe à un temps que si le fait que x existe *simpliciter* obtient, alors le fait que x existe à un temps obtient.

P_2 . Cela fait partie de la nature du fait que x existe *simpliciter* que si le fait que x existe à un temps obtient, alors le fait que x existe *simpliciter* obtient.

Or, ni (P_1) ni (P_2) ne résultent de la remarque de Fine sur l'existence d'un objet à un temps et son existence *simpliciter*. Comme je l'ai dit, son observation concerne les propriétés essentielles d'un objet concret alors que (P_1) et (P_2) concernent les propriétés essentielles de faits. Et Fine aura bien du mal à trouver du soutien indépendant pour (P_1) et (P_2). Cela étant dit, l'échec de principes aussi généraux que (P_1) et (P_2) n'exclut pas qu'une version restreinte de ces principes, par exemple, à propos d'objets existant dans le temps uniquement, soit acceptable. Prenons un exemple particulier :

p_i Cela fait partie de la nature du fait que Socrate existe à un temps que : si le fait que Socrate existe *simpliciter* obtient, alors le fait que Socrate existe à un temps obtient.

p_{ii} Cela fait partie de la nature du fait que Socrate existe *simpliciter* que si le fait que Socrate existe à un temps obtient, alors le fait que Socrate existe *simpliciter* obtient.

Ces deux propositions constitueraient un contre-exemple à (D_{pred}). Nous avons établi que la remarque de Fine ne permettait pas de conclure à de telles propositions par un argument valide. Mais il n'en reste pas moins que p_i et p_{ii} sont plutôt plausibles indépendamment. Je ne compte pas ici développer un argument complet contre p_i ou p_{ii} . Mais une piste pour défendre (D_{pred}) consisterait à rejeter p_i et à dire que malgré que l'implication qu'elle contient soit vraie et thématiquement liée au fait que Socrate existe à un temps, elle ne fait pas partie de la nature du fait en question.

Mon opposant pourrait à ce stade évoquer un principe très plausible sur lequel baser p_i : « [...] pour tout x , si **ce** x existe dans le temps, alors si **ce** x existe *simpliciter*, alors **ce** x existe dans le temps²¹. » Je répondrai simplement que même si le principe en question est indiscutable – et je ne nie pas cela, il n'entraîne nullement p_i , qui est nettement plus substantiel, et je vois mal comment il pourrait le motiver. En effet, le « nouveau » principe est simplement une tautologie de la logique propositionnelle « si p alors (si q alors p) », alors que p_i est un principe substantiel qui concerne l'essence des faits d'existence. La lectrice pourrait aussi penser que si l'on croit comme moi qu'il en va de la nature du fait que la balle est colorée que si le fait que la balle est rouge obtient alors le fait que la balle est colorée obtient également, on accepte forcément p_i . L'idée serait ici que les cas sont similaires et qu'ainsi, le jugement essentialiste devrait également l'être. La différence importante que je souhaite mentionner ici est qu'il y a une connexion de nature entre « être coloré » et « être rouge », puisqu'être coloré c'est essentiellement être « rouge ou vert ou... ». Or, il n'est pas si clair que ce genre de rapport existe entre « exister à un temps » et « exister *simpliciter* ». Du reste, même si ce genre de lien existait, il s'agirait probablement d'un lien asymétrique, comme dans le cas de « coloré » et de « rouge ». En somme, si l'on pense qu'exister *simpliciter* revient essentiellement à exister à un temps ou à exister hors du temps ou... , alors l'analogie entre les deux cas ne permettra pas d'obtenir p_i , mais seulement p_{ii} . Je précise toutefois que mon intention ici n'est pas d'éliminer définitivement l'objection qui repose sur p_i , mais plutôt de montrer que l'attrait intuitif de p_i est sûrement indu, et donc qu'en l'absence de soutien indépendant pour p_i , l'objection ne tient pas.

La quatrième objection concerne la quantification existentielle et la manière dont elle est habituellement fondée. La plupart des théoriciens de la fondation pensent que le fait que quelqu'un est un philosophe est fondé, par exemple, dans le fait que Socrate est un philosophe. Par la deuxième clause de (D_{pred}), cela donne :

P_3 . La conditionnelle suivante fait partie de la nature du fait que quelqu'un est un philosophe : si le fait que Socrate est un philosophe obtient, alors le fait que quelqu'un est un philosophe obtient.

D'après Fine, cette attribution essentialiste n'est pas légitime :

[...] it does not lie in the nature of the fact that someone is a philosopher that the fact is so grounded given that Socrates is indeed a philosopher. The fact, so to speak, knows nothing of Socrates²².

Toutefois, il m'est d'avis que l'appel intuitif de l'objection à (D_{pred}) est plutôt faible. En effet, je vois mal en quoi la connexion essentielle entre un fait et un objet du domaine sur lequel ce fait quantifie serait contre-intuitive. Je comprends tout à fait que Correia se dise « peu impressionné » par l'objection en question. Néanmoins, je pense que la réponse qui suit fera peut-être douter ceux qui partagent l'intuition de Fine à propos du fait que quelqu'un est un philosophe. Prenez par exemple ($f_{iV} \vee f_V$). Il nous semble tout à fait naturel, étant donné la structure logique de ce fait, de dire que cela fait partie de son essence que si f_{iV} obtient, alors ($f_{iV} \vee f_V$) obtient également. Or, une manière tout à fait légitime de comprendre la structure logique de la quantification existentielle présente dans le fait que quelqu'un est un philosophe, c'est-à-dire « $\exists x(Fx)$ », est de l'interpréter comme une grande disjonction contenant tous les objets du domaine ($Fa \vee Fb \vee \dots$)²³. Ainsi, je suggère que la nature du fait qui contient le quantificateur existentiel est une grande disjonction de conditionnelles mentionnant un à un les objets du domaine, comme suit :

Si le fait que Socrate est un philosophe obtient, alors le fait que quelqu'un est un philosophe obtient,

Si le fait qu'Obama est un philosophe obtient, alors le fait que quelqu'un est un philosophe obtient,

...

Une telle conception du fait en question me semble disponible et explique très bien pourquoi le fait que quelqu'un est un philosophe est fondé dans le fait que Socrate est un philosophe. Il y a également une variante de cette objection qui est réglée par un traitement similaire. On dit typiquement que le fait que la balle est colorée est fondé dans le fait que la balle est rouge. Mais, d'après Fine, les faits sur la couleur « ignorent tout » des couleurs particulières. À nouveau, je trouve l'intuition peu mordante, tout particulièrement si l'on comprend la propriété d'être coloré comme une disjonction des différentes couleurs. Je ne vois pas de problème à inclure dans la nature du fait que la balle est colorée :

Si le fait que la balle est rouge obtient, alors le fait que la balle est colorée obtient,

Si le fait que la balle est bleue obtient, alors le fait que la balle est colorée obtient,

...

LES OBJECTIONS « CONTINGENTISTES »

La classe d'objections que je considérerai dans cette section n'est pas évoquée par Correia. Elle provient du débat sur le rapport entre la fondation et la nécessité. Une conséquence peu surprenante de la définition de la fondation en terme d'essence est la thèse habituellement appelée « nécessitarisme de la fondation » :

N_{fon} Si le fait que p est fondé dans q_i, q_{ij}, \dots , alors nécessairement²⁴ ($q_i, q_{ij}, \dots \rightarrow p$).

En effet, si les rapports de fondation sont issus de la nature des faits, il y a fort à parier que ces rapports tiendront nécessairement. Par conséquent, s'il y a des cas de fondation contingente, ceux-ci constituent *ipso facto* des contre-exemples directs à (D_{pred}). Les cas les plus convaincants de fondation contingente sont les généralisations accidentelles²⁵. Prenons le fait que le membre le plus grand du département de philosophie de Genève fait moins de deux mètres, que j'abrégerai en f_I . Il semble naturel de dire que :

G_{acc} f_I est fondé dans les faits individuels f_i, f_{ij}, \dots qui concernent la taille respective de chacun des membres du département de philosophie de Genève.

Imaginons maintenant un monde possible, fut-il surprenant, dans lequel Shaquille O'Neal est un membre du département de philosophie de Genève. Les faits individuels f_i, f_{ij}, \dots qui fondaient f_I obtiennent dans ce monde, mais f_I n'y obtient pas, car O'Neal mesure deux mètres seize. Il y a donc quelqu'un de plus grand que deux mètres au département, *contra* f_I . Notre définition avec G_{acc} implique qu'il est dans la nature de f_I que si f_i, f_{ij}, \dots obtiennent, alors f_I . Or, j'ai montré que ces derniers pouvaient obtenir sans que f_I n'obtienne. Il est aisé de voir que d'innombrables objections du même type peuvent être générées, ce qui rend d'autant plus pressant le besoin d'y répondre. Voici donc une manière dont on peut répondre à l'objection contingentiste : les fondements cités dans G_{acc} sont tout bonnement incomplets : il s'agit de *fondements partiels*. La notion de fondation que l'on cherche à définir fournit quant à elle des explications complètes²⁶. Ainsi, puisqu'il manque quelque chose aux faits f_i, f_{ij}, \dots qui concernent la taille respective de chacun des membres du département de philosophie de Genève pour expliquer pleinement f_I , la collection de faits f_i, f_{ij}, \dots ne fonde pas f_I . La structure de l'argument est assez simple :

Prémisse 1 : Si x fonde y , alors x explique complètement y .

Prémisse 2 : f_i, f_{ij}, \dots n'expliquent pas pleinement f_I .

Conclusion : f_i, f_{ij}, \dots ne fonde pas f_I .

On pourrait se montrer sceptique à l'idée selon laquelle G_{acc} est seulement une explication partielle. De fait, dans les cas d'explication partielle, on échoue typi-

quement à rendre compte d'une partie de l'*explanandum*. Supposons que je veuille expliquer le fait que Kit Fine est un philosophe anglais et que je cite le fait que Fine est un philosophe. Il est assez clair que je n'ai donné qu'une explication partielle. Une manière de le voir est de constater que je n'ai pas rendu compte du fait que Fine est Anglais. Pour revenir à G_{acc} , on pourrait répondre à l'idée qu'il s'agit de fondement partiel qu'il n'y a aucune partie de l'*explanandum* dont on n'a pas rendu compte. Je pense que cette réponse est naïve, car il y a clairement une partie du contenu informatif de f_I dont on n'a pas rendu compte. Supposons que je demande à un enfant : « Quel est l'enfant le plus grand de ta classe ? » et que ce dernier me réponde : « Je mesure un mètre trente, Sylvie mesure un mètre trente-trois et Pauline un mètre quarante. » Je serai naturellement amené à lui demander s'il s'agit là de tous les membres de la classe. S'il répond « oui », je disposerai alors d'une explication complète. En somme, et pour reprendre l'exemple qui nous intéresse, la partie dont on n'a pas rendu compte est l'idée ou le fait que les gens listés dans f_i, f_{ij} ... sont tous les membres du département qu'il y a. Une fois ce fait ajouté aux fondements de f_I , l'ajout d'O'Neal à la liste des membres du département, si surprenante soit-il, ne constituera plus un contre-exemple. En effet, une partie des faits qui fondaient f_I n'obtiendront plus, car la liste des membres du département aura changé.

Un problème notoire de cette réponse est qu'elle demande de postuler une quantité massive de faits de totalité, comme le fait qu'untel et untel sont tous les membres du département de philosophie de Genève, le fait qu' a, b et c sont toutes les chaises dans cette salle, le fait qu' a_i, a_{ij} ... sont tous les atomes qui composent cette chaise, etc. Pire encore, bon nombre de ces faits seront fondamentaux, dans le sens où ils ne seront pas fondés dans d'autres faits. Ces considérations sont particulièrement gênantes si l'on souscrit à un principe similaire au *purity principle* de Sider²⁷, qui veut que seules des entités fondamentales soient mentionnées dans les faits fondamentaux²⁸. Pour les philosophes qui ont un projet d'ontologie du fondamental, la conséquence selon laquelle des entités telles que le département de philosophie de Genève, certaines chaises ou même une foule de faits généraux à propos des choses qu'il y a dans tel et tel lieux soient des entités fondamentales est particulièrement dévastatrice. Ainsi, la solution des faits de totalité semble mettre en danger le projet même d'ontologie néo-aristotélicienne²⁹, selon lequel ce qui est important en ontologie n'est pas de déterminer les choses qui existent, mais plutôt de déterminer quelles sont les entités fondamentales de notre monde³⁰.

Il y a deux parties à ce problème, la première étant que certains faits s'avèrent fondamentaux et viennent gonfler l'ontologie du fondamental. Toutefois, cet aspect ne semble pas concerner les opérationnistes qui traitent la fondation comme un opérateur lequel utilise au moins deux phrases pour faire une phrase. Ces derniers ne sont donc pas engagés envers une ontologie des faits. Par contre, il leur reste toujours une quantité de phrases générales inexplicées, comme « untel et untel sont tous les membres du département de philosophie de Genève », « a, b et c sont toutes les chaises dans cette salle », « a_i, a_{ij} ... sont tous les atomes qui composent cette chaise. » Ce qui, en admettant un *purity principle* qui

concerne les phrases, implique quand même une partie des conséquences indésirables mentionnées plus haut. En particulier, l'inclusion d'entités comme le département de philosophie dans les entités fondamentales. Je préconise donc de rejeter ce genre de principe, que j'estime par ailleurs³¹ trop inclusif. Il y a d'autres manières de dériver une ontologie du fondamental à partir de la fondation. Par exemple, on peut adopter le principe selon lequel, si le fait que x existe est fondé dans le fait qu' y existe, alors y est plus fondamental que x . Un tel principe n'est pas une définition de la notion de fondamentalité, mais il nous permet de savoir quels sont les objets fondamentaux, pourvu que l'on sache quels faits d'existence sont fondés dans quels autres faits d'existence. Ce genre de principe a l'avantage de n'être sensible qu'à une certaine classe de phrases. Typiquement, il ne prendra pas en compte des phrases de la forme « $a=a$ ». Par ailleurs, une fois dérivés les rapports d'« être plus fondamental que », il est aisé d'en obtenir la converse « être moins fondamental que », puis d'obtenir par la négation de ces deux le rapport d'« être aussi fondamental que ». En effet, si x est plus fondamental que y , alors y est moins fondamental que x , et si x n'est ni plus ni moins fondamental que y , alors x est « aussi fondamental que » y et vice-versa. Je n'en dirai pas plus à ce sujet, car il n'est pas dans la portée de cette article de montrer comment la définition de la fondation que j'ai présentée ici permet d'accomplir le projet général d'une ontologie du fondamental. L'objet de ce dernier passage était seulement de montrer qu'il y a des alternatives disponibles au modèle siederéen et que, du coup, la solution au « contingentisme » peut éviter certaines conséquences négatives.

Par ailleurs, j'aimerais préciser que j'ai développé mes arguments en faveur d'une définition qui présente la fondation comme un prédicat, cela, principalement pour des raisons de clarté. Une version analogue de cette définition exploitant un opérateur permet d'y répondre aussi bien, quoiqu'un peu différemment. Une telle définition ressemblerait à :

S_{op} p parce que p_i, p_{ii}, \dots

ssi (déf.)

i) $p_i \& p_{ii} \& \dots$, et

ii) La conditionnelle suivante fait partie de ce que cela fait qu'il soit le cas que p : si $p_i \& p_{ii} \& \dots$, alors p .

Pour des raisons de simplicité et de clarté, j'ai adopté la notation des faits et le langage prédicationniste. C'est du reste aussi le cas de Fine dans son « Guide to Ground » qui exploite les faits pour simplifier l'exposition, bien qu'il soit officiellement opérationniste. C'est une curiosité à noter qu'à ce jour, la plupart des théoriciens qui pensent que la fondation doit *in fine* être exprimée à l'aide d'un opérateur utilisent quand même l'approche prédicationnelle à des fins d'exposition.

CONCLUSION

J'espère avoir montré que l'orthodoxie selon laquelle la notion de fondation doit être adoptée comme primitive et ne peut être définie est une erreur. On peut répondre de manière concluante à toutes les objections à (D_{pred}) avancées par Kit Fine. Par ailleurs, j'ai mobilisé de nouvelles objections provenant du débat sur le nécessitarisme à propos de la fondation, après quoi j'ai proposé des solutions à ces nouvelles objections qui permettent de garder la définition intacte. Il en résulte que les rapports entre fondation et essence sont peut-être bien plus étroits que Fine ne le pense et que la notion de fondation n'est, après tout, pas primitive.

Au-delà d'une compréhension plus claire des rapports entre fondation et essence, (D_{pred}) est un outil idéal de parcimonie conceptuelle. Plutôt que deux concepts primitifs, pour ceux qui adoptaient l'essence et la fondation dans leur théorie, il n'en faut plus qu'un. En outre, cela peut également convaincre ceux qui voyaient dans la fondation une relation obscure qu'elle peut être comprise, pourvu qu'on ne soit pas également sceptique à propos de la notion d'essence.

NOTES

- ¹ Voir Fine, Kit, "Essence and Modality," *Philosophical perspectives*, vol. 8, 1994, pp. 1-16.
- ² J'utiliserai ce terme de manière artificielle pour désigner plus étroitement la notion de fondation *métaphysique*. On peut penser qu'il y a aussi une notion de fondation scientifique, et même normative; c'est le cas notamment de Fine. Mon propos dans cet article se concentre sur la notion métaphysique.
- ³ Ce connecteur doit être compris dans un sens non causal. Ainsi je dirai que Sam est triste en vertu de son état neurophysiologique, mais pas qu'il est triste en vertu de son récent échec à l'examen. Parallèlement, je dirai que le fait que Sam est triste est fondé dans certains faits à propos de l'état neurophysiologique de Sam, mais pas que le fait que Sam est triste est fondé dans le fait qu'il a échoué à l'examen.
- ⁴ Fine, Kit, "Guide to Ground," in Correia, Fabrice et Benjamin Schnieder (eds.), *Metaphysical Grounding: Under-Understanding the Structure of Reality*, Cambridge, Cambridge University Press, 2012, p. 37; Rosen, Gideon, "Metaphysical Dependence: Grounding and Reduction," in Hale, Bob et Aviv Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology*, Oxford, Oxford University Press, 2010, pp. 109-136; Schaffer, Jonathan, *Grounding as the Primitive Concept of Metaphysical Structure*, manuscrit.
- ⁵ Correia, Fabrice, "Metaphysical Grounds and Essence," in Miguel, Hoeltje, Schnieder, Benjamin et Alexander Steinberg (eds.), *Varieties of Dependence. Ontological Dependence, Grounding, Supervenience, Response-Dependence*, Munich, Basic Philosophical Concepts Series, Philosophia, 2013, pp. 271-296.
- ⁶ Fine, Kit, "Guide to Ground", *op. cit.*, p. 37.
- ⁷ Audi, Paul, "Grounding: Toward a Theory of the In-Virtue-Of Relation," *Journal of Philosophy*, vol. 109, no. 12, 2012, pp. 685-711; Dasgupta, Shamik, "The Status of Ground," (manuscrit); Fine, Kit, "Guide to Ground," *op. cit.*; Dorr, Cian, "To be an F is to be a G," (manuscrit).
- ⁸ Quand je parle d'un « fait », j'entends simplement la portion du monde qui correspond à une proposition vraie. Ainsi, en admettant que Sam soit triste, la proposition « Sam est triste » est vraie et il y a une partie du monde qui, en quelque sorte, « correspond » à la vérité de cette proposition. J'appelle cette partie du monde *le fait que Sam est triste*. J'entends ici rester neutre sur au moins deux questions importantes concernant les faits : la première est de savoir quelle est leur structure logique, et, notamment, s'il y a des faits négatifs. La seconde, non sans lien avec la première, est de savoir si deux faits strictement co-extensionnels peuvent être différents. Par exemple, le fait que Sam est célibataire et le fait que Sam n'est pas marié. L'intuition selon laquelle Sam est célibataire *parce qu'il n'est pas marié* suggère que ces deux faits ne sont pas les mêmes, puisque le connecteur explicatif « parce que » est irréflexif. Ces questions, et bien d'autres à propos des faits, se trouvent hors de la portée du présent argument.
- ⁹ Correia, Fabrice, "Grounding and Truth-Functions," *Logique et Analyse*, vol. 53, no. 211, 2010, pp. 251-279.
- ¹⁰ Pour davantage de considérations sur la différence entre opérationisme et prédicationisme, voir Correia et Schnieder, "Grounding: An Opinionated Introduction," *op. cit.*, p. 1.
- ¹¹ Cette clause présume d'une compréhension de la fondation comme *factive*, c'est-à-dire telle que si *f* fonde *g*, alors *f* obtient et *g* obtient. Cette conception est généralement acceptée, mais comme le signale une relectrice anonyme, on peut aussi imaginer une relation de fondation non factive connectant, parfois, des états de choses n'obtenant pas. Une manière simple de définir une telle notion de fondation est de retirer la première clause du *definiens*, puisque cette clause est là pour garantir la factivité de la relation en question. Il faudrait à ce moment-là que l'on dispose d'une notion adéquate d'essence pour des faits ou des états de choses qui n'obtiennent pas, ou que l'on traite les états de choses n'obtenant pas comme des objets au même titre que les faits. Ainsi, la non-factivité requiert quelques modifications substantielles non sur la forme de la définition, mais sur la métaphysique qu'elle exploite. Par contre, une fois ces

modifications fournies, les résultats de la définition seront sensiblement similaires relativement aux objections que j'aborde ici. La notion factive de fondation reste plus intéressante à définir puisqu'elle est très majoritairement utilisée (Fine, malgré qu'il suggère l'idée d'une notion non factive, exploite principalement la notion factive). Je remercie au passage la relectrice anonyme d'avoir mentionné ce point intéressant.

- ¹² J'ai présenté les faits comme s'ils correspondaient aux propositions vraies. La notion d'état de choses, quant à elle, est plus large, car ces états peuvent également correspondre à des propositions fausses. On peut donc distinguer les états de choses qui n'obtiennent pas, par exemple, l'état de choses selon lequel la Suisse est en Afrique, des états de choses qui obtiennent, par exemple l'état de choses selon lequel la Suisse est un pays. Une fois que la distinction est introduite, on dira, typiquement, que les faits sont des états de choses qui obtiennent. Il y aurait également beaucoup à dire ici, mais cette caractérisation extrêmement simpliste devrait suffire pour les besoins de cet article.
- ¹³ Mes remerciements vont à Fabrice Correia pour m'avoir fait remarquer, lors de communications personnelles, que la notion d'essence d'un fait n'est pas une nouvelle notion d'essence. Il s'agit simplement de l'essence d'un objet (ou essence objectuelle) prenant pour objet un fait.
- ¹⁴ Il faut toutefois noter que l'intuition essentialiste n'est pas libre de controverses. Des philosophes aussi importants que Spinoza, Leibnitz, Quine et Lewis l'ont historiquement rejetée. Mes remerciements vont à une relectrice anonyme pour avoir proposé cette remarque.
- ¹⁵ Fine, Kit, "Guide to Ground", *op. cit.*, p. 37.
- ¹⁶ Par « fondement », j'entends ce qui fonde plutôt que la relation, que j'appelle « fondation ».
- ¹⁷ Le principe en question veut que si x fonde y et qu' y fonde x , alors y n'est pas identique à x . Autrement dit, il n'y a pas d'instance de x fonde x . Ce principe est accepté par la vaste majorité des théoriciens de la fondation, Fine inclus.
- ¹⁸ Je précise ici que la caractérisation d'essence constitutive que je propose, quoi qu'inspirée de Fine, n'est pas sa définition officielle. Fine définit l'essence constitutive en terme d'essence conséquentielle. Un défaut d'une telle approche est qu'elle semble inverser l'ordre de la compréhension; nous avons typiquement un accès intuitif à l'essence de Socrate et, à partir de cela, nous pouvons dériver une classe de propositions plus large le concernant. Mais nous n'avons aucun accès intuitif à l'essence conséquentielle de Socrate. Pour une critique plus complète de la définition d'essence constitutive en terme d'essence conséquentielle, voir Fabrice Correia, "More on the Reduction of Necessity to Essence", manuscrit.
- ¹⁹ Fine, Kit, "Senses of Essence," in W. Sinnott-Armstrong (ed.), *Modality, Morality and Belief*, Cambridge, Cambridge University Press, 1995, p. 57.
- ²⁰ Fine, "Guide to Ground", *op. cit.*, p. 79.
- ²¹ Mes remerciements vont à une relectrice anonyme pour avoir insisté sur l'importance de cette objection et sur le soutien dont elle bénéficie.
- ²² *Ibid.*, p. 75.
- ²³ C'est ainsi, par exemple, que Wittgenstein conçoit la nature d'une telle proposition. Wittgenstein, Ludwig, *Tractatus logico philosophicus*, Londres, Routledge et Kegan Paul, 1974.
- ²⁴ Nécessité métaphysique.
- ²⁵ Ces cas sont discutés dans l'article d'Alexander Skiles, "Against Grounding Necessitarianism" (manuscrit), où il présente également des cas de la littérature sur la composition matérielle comme le « sandwich de Thésée ».
- ²⁶ D'aucuns pourront penser qu'il s'agit là d'une pétition de principe à l'encontre des contingentistes, mais il n'en est rien. Le contingentiste peut tout à fait soutenir de manière cohérente la thèse selon laquelle la fondation fournit des explications complètes, mais que celle-ci n'a pas la force modale que leur attribuent les nécessitaristes. Dans cet esprit, le contingentiste tâchera de nier la deuxième prémisse de l'argument qui suit.
- ²⁷ Sider, Theodore, *Writing the Book of the World*, Oxford, Oxford University Press, 2011.
- ²⁸ Un fait fondamental est simplement un fait non fondé.

²⁹ Schaffer, Jonathan, “On What Grounds What,” in Manley, David, Chalmers, David et Ryan Wasserman (ed.), *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford, Oxford University Press, 2009, p. 347.

³⁰ Je tiens ici à rester neutre vis-à-vis d’un tel projet. Ma remarque est simplement que le projet est en conflit avec les faits de totalité postulés par ma réponse au défi contingentiste.

³¹ Pour ne mentionner qu’un contre-exemple : à supposer que les identités réflexives ($a=a$) ne soient pas explicables, on devrait alors accepter toute entité comme fondamentale.

SEMANTIQUE FORMELLE ET ENGAGEMENT ONTOLOGIQUE ¹

THIBAUT GIRAUD

DOCTORANT À L'INSTITUT JEAN NICOD, EHES

RÉSUMÉ :

Je montrerai en premier lieu comment et pourquoi la sémantique formelle peut être employée comme un outil pour déterminer l'engagement ontologique d'une théorie : je soutiendrai d'une part que la sémantique doit être prise au sérieux comme apte à décrire la vérification des formules du langage; d'autre part, que les engagements ontologiques d'une théorie sont déterminés par ses vérificateurs. De là, j'exposerai une méthode générale permettant, étant donné un certain type d'ontologie, de construire une sémantique dont les engagements ontologiques sont en accord avec celle-ci. Pour cela, je définirai la notion de *cadre ontologique* : il s'agit d'une structure telle que toute sémantique construite à partir de cette structure aura un certain engagement ontologique déterminé à l'avance. J'exposerai quatre cadres représentant deux types de nominalisme et deux types de réalisme, et j'esquisserai à partir de ces cadres quatre sémantiques pour les langages du premier ordre.

ABSTRACT:

In a first part, I will show why formal semantics can be used as a tool to determine the ontological commitment of a theory: first I will argue that we should take semantics seriously as describing the truthmaking of formulas of language, and then I will claim that the ontological commitments of a theory are determined by its truthmakers. From there, I will set forth a general method allowing us, for a given kind of ontology, to construct a semantics whose ontological commitments are in accordance with this ontology. For that purpose, I will define the notion of *ontological framework*: it is a set-theoretical structure such that any semantics constructed from this structure will bear certain predetermined ontological commitments. I will expose four ontological frameworks standing for two kinds of nominalism and two kinds of realism, and I will construct from those frameworks four distinct semantics for first-order languages.

LA SEMANTIQUE FORMELLE : UN GUIDE POUR L'ENGAGEMENT ONTOLOGIQUE

La sémantique formelle décrit une structure de vérification

Qu'est-ce qu'une sémantique formelle pour un langage \mathcal{L} envisagée du strict point de vue logique? Cela consiste en deux choses : (i) la définition de ce qu'est un modèle pour \mathcal{L} ; (ii) des règles sémantiques pour interpréter les formules de \mathcal{L} dans n'importe quel modèle.

(i) Un modèle est généralement défini comme un n -uplet satisfaisant certaines conditions. Par exemple, pour la sémantique standard de la logique du premier ordre, un modèle est un couple $\langle \mathcal{D}, \delta \rangle$ tel que \mathcal{D} est un ensemble non vide, et que δ est une fonction associant à chaque constante individuelle un membre de \mathcal{D} et à chaque prédicat n -adique un ensemble de n -uplets de membres de \mathcal{D} . Toute structure ensembliste qui satisfait les conditions prescrites est de fait un modèle pour le langage \mathcal{L} .

(ii) On formule un ensemble de règles sémantiques qui déterminent l'attribution d'une (ou de plusieurs) valeur(s) aux formules de \mathcal{L} dans n'importe quel modèle donné. Dans le cas le plus simple, il y a une seule valeur généralement appelée *vérité* ou *vrai* – mais on pourrait tout aussi bien considérer que cette valeur est 1 ou n'importe quoi d'autre; le nom que l'on donne à cette valeur ne change rien du point de vue logique.

Il n'est pas évident de voir en quoi la sémantique formelle serait une source d'information pertinente pour l'ontologie. On pourrait soutenir qu'une sémantique n'est rien de plus qu'un instrument formel servant à prouver certaines propriétés d'un système déductif; c'est une attitude vis-à-vis de la sémantique adoptée par Edward Zalta, par exemple :

It is important to remember that the formal semantics simply provides a set-theoretical framework in which models of the metaphysical theory may be constructed. The models serve the heuristic purpose of helping us to visualize or picture the theory in a rigorous way. It is extremely important not to confuse the models of the theory with the world itself. Many theorists today tend to models of a theory an exalted status that they do not have. [...] As far as the present work is concerned, all that the models of a theory do is show that the theory is consistent, that the logic is complete, that the axioms are categorical, and so forth. (Zalta, 1988, pp. 34-35)

La sémantique telle que Zalta la conçoit est un simple outil logique sans aucune signification ontologique. Un modèle ne représente en aucun sens un monde, et la notion définie de *vérité dans un modèle* n'a rien à voir avec la notion métaphysique de vérité. Cette conception modeste de la sémantique est tout à fait tenable, mais je voudrais expliquer maintenant pourquoi il me semble que l'on a de bonnes raisons de faire jouer à la sémantique un rôle plus intéressant.

Je suis réaliste (anti-idéaliste) au sens où j'admets les présupposés suivants :

- Il existe un monde indépendant de notre langage.
- La vérité d'au moins certaines propositions de notre langage dépend de certains aspects du monde.

- Il y a donc une relation monde/langage qui est une relation vérificative : le monde (la façon dont certaines choses sont dans les mondes) rend vraies certaines propositions de notre langage.

Si l'on accepte ces présupposés réalistes, l'idée que la sémantique pourrait être employée pour décrire la structure de cette relation monde/langage devient très séduisante : un modèle pourrait servir à représenter la structure d'un monde, et les règles sémantiques permettraient alors de décrire comment un monde ayant une certaine structure rend vraies telles et telles propositions du langage. En un mot : si l'on est réaliste à l'égard de la vérité, il semble naturel d'attendre de la sémantique d'un langage qu'elle serve à l'étude de la vérification de ce langage. Utiliser la sémantique dans ce dessein, c'est ce que j'appellerai à partir de maintenant *prendre la sémantique au sérieux*.

Toutefois, l'idée qu'un modèle représente la structure d'un monde peut dérouter, parce que dans la sémantique de la logique modale (souvent appelée « sémantique des mondes possibles »), certains éléments d'un modèle sont supposés représenter des « mondes » ou des « mondes possibles », aussi semblerait-il inapproprié de penser à un tel modèle comme s'il était apte à représenter la structure d'un *monde*. On pourrait alors préférer dire que ce modèle représente la structure d'un *univers* de mondes possibles; ainsi les règles sémantiques décriraient comment cet univers modal rend vraies ou fausses les formules du langage modal. On peut proposer une option encore plus neutre pour éviter de parler spécifiquement de monde aussi bien que d'univers : on pourrait dire qu'un modèle représente la structure d'une *situation ontologique*, c'est-à-dire n'importe quoi pouvant tenir lieu de réponse à la question ontologique : qu'est-ce qu'il y a?

Pour conclure, si l'on prend la sémantique au sérieux, c'est-à-dire si on la considère comme apte à décrire la relation monde/langage, alors définir ce qu'est un modèle revient à donner la structure d'une situation ontologique en général, et les règles sémantiques peuvent être considérées comme des règles décrivant, pour n'importe quelle situation ontologique et n'importe quelle formule du langage, comment cette situation ontologique rend vraie ou non cette formule; en un mot, les règles sémantiques décrivent la vérification des formules.

La vérification détermine l'engagement ontologique

Je défendrai la thèse selon laquelle les engagements ontologiques d'une phrase ou d'une théorie sont déterminés par ce qui *rendrait vraie* cette phrase ou cette théorie. Un tel point de vue est défendu par divers auteurs (Simons, 1997; Heil, 2003; Armstrong, 2004; Cameron, 2008; 2010). Certes, nous offensoons par là l'orthodoxie quiniennne sur l'engagement ontologique, mais on peut noter que de nombreux philosophes ont déjà solidement argumenté contre elle (on peut lire par exemple l'excellente critique de Prior, 1971, ch. 3, à propos de la quantification de second ordre), et il me semble que, dès lors que l'on est réceptif à l'idée de vérification, l'idée que celle-ci détermine l'engagement ontologique devient tout à fait naturelle et intuitive, et constitue même, en un certain sens, une approche plus modeste que l'approche de Quine.

Ce passage de Simons décrit très précisément comment les deux notions de vérification et d'engagement ontologique semblent liées :

Ontological commitment is a sort of converse to an idea which is of more recent prominence: truth-making. Whereas when we ask what things are such that their existence is *necessary* for a sentence to be true, we are asking after its ontological commitments; when we ask what things are such that their existence is *sufficient* for the sentence to be true, we are considering the sentence's *truth-makers*. [...] We could therefore characterize the ontological commitments of a sentence or sentences as given by the least that would be required to make it true. (Simons, 1997, p. 265)

Cette dernière phrase exprime assez clairement ce qu'on pourrait appeler l'approche vérificationnelle de l'engagement ontologique. L'approche quantificationnelle (c'est-à-dire l'approche standard de Quine) et l'approche vérificationnelle ne s'opposeraient pas si le vérificateur d'une phrase existentiellement quantifiée comme « il y a un x » était toujours l'existence de cet x . Mais le défenseur de l'approche vérificationnelle peut précisément rester neutre sur ce point ou le rejeter. Comme Cameron le souligne :

I think one of the benefits of truthmaker theory is to allow that $\langle x$ exists \rangle might be made true by something other than x , and hence that “ a exists” might be true according to some theory without being an ontological commitment of that theory. (Cameron, 2008, p. 401)

Développons ce point à partir d'un exemple. Selon le critère quinién, si la phrase « Il y a une table » est littéralement vraie, nous sommes ontologiquement engagés envers l'existence d'une table; la table doit être une entité distincte dans notre ontologie. Aussi, si nous pensons qu'une table n'est en fait qu'un tas d'atomes arrangés de telle et telle façon, nous ne devrions pas prendre « Il y a une table » pour une vérité littérale, mais comme une manière impropre de dire « Il y a tels et tels atomes arrangés de telle et telle façon ». À l'inverse, si nous adoptons le critère de la vérification, « Il y a une table » n'engage ontologiquement qu'à *ce qui rend vraie cette phrase*, quoi que ce soit. Plus précisément, il faut en fait considérer un vérificateur *minimal*. Si nous supposons qu'un vérificateur (minimal) pour « Il y a une table » consiste dans le fait qu'un tas d'atomes sont arrangés de telle et telle façon, alors cette phrase n'engage ontologiquement qu'en l'existence d'atomes arrangés de cette façon. Ainsi, on peut considérer que la phrase « Il y a une table » est littéralement vraie sans que cela engage ontologiquement l'existence d'une nouvelle entité en plus des atomes qui la composent.

Un autre exemple particulièrement éclairant est donné par Simons :

This way of looking at ontological commitment highlights something which might otherwise remain clouded, and which one might call the inscrutability of ontological commitment. Consider first a simple medical sentence such as *Sam has hepatitis*. This is made true by hepatitis viruses in sufficient numbers in Sam's body, causing inflammation of his liver. But we cannot tell this by simply analyzing the sentence linguistically: it is a matter of medical knowledge, not conceptual analysis. Not even the type of virus is fixed by the statement: when in a paper on truth-making, Mulligan *et al.* put forward this example, two types of hepatitis virus were known, since then another has come to light, which only serves to underline the point that it is in general an *a pos-*

teriori matter what makes a given sentence true. But if that is so then the ontological commitments or truth-making minima of a sentence are not to be read off its logico-grammatical form either: when we affirm that Sam has hepatitis then all we are committed to is *something or other, whatever it is* that causes Sam's liver to be inflamed. (Simons, 1997, p. 265)

Bien sûr, cette approche de l'engagement ontologique n'est pas exempte de difficultés, mais il me semble qu'il s'agit bien là de la meilleure manière de traiter cette question : il est clair qu'il y a un lien très étroit entre *ce qui rend vraie* une théorie, c'est-à-dire ses vérifacteurs, et *ce qui est* d'après cette théorie, c'est-à-dire ses engagements ontologiques. On peut notamment remarquer que sises vérifacteurs existent, alors la théorie est vraie, donc ses engagements ontologiques existent eux aussi. Comprendre la vérification d'une théorie permet de bien déterminer ce à quoi elle engage ontologiquement.

À partir de cette approche vérificationnelle de l'engagement ontologique, si l'on prend maintenant la sémantique formelle au sérieux, c'est-à-dire si l'on considère qu'elle décrit la vérification des formules du langage dans n'importe quelle situation ontologique, alors il semble que la sémantique formelle puisse nous apporter des informations précises quant à l'engagement ontologique du langage interprété par la sémantique. Cela revient-il à dire que nous devons laisser la sémantique nous dicter notre ontologie? Non. L'idée est plutôt de procéder dans l'autre sens. Dès lors que nous comprenons bien la relation entre la sémantique formelle et l'ontologie, il devient possible de construire une sémantique de telle façon que, étant donné un type d'ontologie, les engagements ontologiques de cette sémantique soient *in fine* en accord avec ce type d'ontologie. Autrement dit, il s'agit de construire la sémantique dans une perspective ontologique.

QUATRE CADRES ONTOLOGIQUES

Je vais présenter une méthode pour construire des sémantiques dont l'engagement ontologique correspondra à certains types classiques d'ontologie. À cette fin, je définirai des *cadres ontologiques*. Pour expliquer en quoi consiste un cadre ontologique, il nous faut nous pencher d'un peu plus près sur la structure d'un modèle en sémantique.

Étant donné un langage \mathcal{L} construit à partir d'un ensemble de constantes non-logiques \mathcal{V} (ce qu'on appelle son *vocabulaire* ou sa *signature*), un modèle pour \mathcal{L} dans une sémantique aura généralement la forme suivante : $\langle X, Y, \dots, \delta \rangle$ où X, Y, \dots sont toutes sortes d'ensembles ou de fonctions, et où δ est une fonction de dénotation, c'est-à-dire qu'elle associe les éléments de \mathcal{V} à d'autres éléments du modèles. (Il pourrait aussi y avoir plusieurs fonctions de dénotation, c'est-à-dire plusieurs fonctions associant des éléments de \mathcal{V} à d'autres éléments du modèle.) En dehors de cette fonction de dénotation qui assure le lien entre le langage et le monde, tous les autres éléments du modèle peuvent être considérés comme des éléments purement ontologiques, qui n'ont pas trait au langage interprété. Aussi, ce que je vais appeler le *cadre ontologique* à partir duquel la sémantique est construite est la structure $\langle X, Y, \dots \rangle$, c'est-à-dire la structure d'un modèle sans la fonction de dénotation. Donc, pour construire une sémantique à partir d'un cadre ontologique, il suffit d'ajouter à la structure du cadre une fonction de dénotation.

Le cadre ontologique, si la sémantique est prise au sérieux, devrait représenter la structure de *ce qu'il y a* en général, la structure de n'importe quelle situation ontologique. Aussi, le choix du cadre devrait être le reflet de nos intuitions ontologiques, indépendamment du langage que l'on souhaite ensuite interpréter dans ce cadre. Cet aperçu général deviendra plus clair à partir de quelques exemples. Je vais considérer quatre types classiques d'ontologie concernant le statut des individus et des universaux. *Nominalisme fort* : le monde est fait seulement d'individus. Tout découpage de la réalité est arbitraire, et tout prédicat n'est qu'un nom qui ne correspond pas à une structure réelle. *Nominalisme modéré* : le monde est fait seulement d'individus, mais tout découpage de la réalité n'est pas arbitraire. Il y a une structure réelle (par exemple de ressemblance) sur les individus. *Réalisme modéré* : ce qui est fondamental, ce sont des états de choses de forme : *un universel n-adique instancié par n individus*. Il y a donc des individus et des universaux, mais ils ne peuvent pas exister indépendamment : ils sont toujours connectés dans un état de choses. *Réalisme fort* : il y a des universaux et leur existence est indépendante de l'existence d'individus les instanciant. Par contre, les individus ne peuvent pas exister sans instancier des universaux.

Le réalisme modéré correspond à une ontologie de type aristotélicien ou armstrongien, et le réalisme fort correspond à un platonisme à l'égard des universaux. (Notons bien que j'utilise maintenant le terme *réalisme* en un autre sens que précédemment : je parlais plus haut du réalisme à l'égard de la vérité, ce que l'on peut opposer à l'idéalisme; maintenant je parle du réalisme *des universaux*, qu'on oppose généralement au nominalisme).

Je vais maintenant présenter quatre cadres ontologiques représentant ces différents points de vue, et j'esquisserai ensuite, à partir de chacun d'eux, la construction d'une sémantique pour les langages du premier ordre.

Cadre nominaliste fort

Le cadre nominaliste fort est le plus simple. Le monde n'est fait que d'individus et il n'y a aucune sorte de structure sur ces individus; ce cadre sera donc simplement une structure $\langle \mathcal{I} \rangle$ où \mathcal{I} est un ensemble d'urelements (qui correspondent intuitivement aux individus).

Je ne veux pas dire que d'après le nominalisme fort, le monde est un ensemble (ne serait-ce que parce qu'il est clair qu'un nominaliste pourrait questionner l'idée qu'il y ait des ensembles). Je dis plutôt que la *structure du monde*, tel qu'un nominaliste fort l'envisage, est adéquatement décrite par la structure $\langle \mathcal{I} \rangle$, c'est-à-dire par un simple ensemble d'individus. Le cadre ontologique représente la *structure* générale de la réalité; il ne faut pas l'identifier à la réalité elle-même.

Afin d'obtenir une sémantique pour un langage à partir d'un cadre ontologique, il suffit d'ajouter au cadre une fonction de dénotation pour ce langage. Une sémantique nominaliste forte devrait donc avoir la structure suivante : $\langle \mathcal{I}, \delta \rangle$. En sémantique standard de la logique du premier ordre, les modèles ont précisément cette structure : δ est une fonction de dénotation associant les constantes individuelles à des membres de \mathcal{I} , et les prédicats à des ensembles de n -uplets de membres de \mathcal{I} . Par exemple, la formule Pa (où P est un prédicat monadique et a une constante individuelle) est vraie ssi $\delta(a) \in \delta(P)$; et la formule Rab est vraie ssi $\langle \delta(a), \delta(b) \rangle \in \delta(R)$.

Ainsi donc, si l'on prend la sémantique de la logique du premier ordre au sérieux, l'engagement ontologique de cette sémantique correspond à un nominalisme fort.

J'appellerai les urelements d'un cadre les *entités* du cadre. Dans le cadre nominaliste fort, les individus sont les seules entités. En un certain sens, on peut considérer les entités comme les *blocs de construction du monde*.

Cadre nominaliste modéré

Dans le cadre nominaliste modéré, il n'y a encore que des individus, mais ceux-ci présentent une certaine structure; admettons qu'ils sont structurés par une relation de ressemblance. (C'est l'approche la plus courante. On pourrait aussi considérer l'idée de classe naturelle; je pense que le cadre obtenu à partir de cette idée ne différerait pas beaucoup de celui que je vais construire ici.) Un simple ensemble \mathcal{I} d'individus ne suffira donc pas à représenter le nominalisme modéré : il faut pouvoir en plus exprimer cette idée de structure de ressemblance.

Intuitivement, en plus d'un ensemble d'individus, on aura besoin d'ensembles qui représentent des individus se ressemblant sous un certain aspect. Par exemple, on aura un ensemble d'individus se ressemblant du point de vue leur rougeur (des pommes, des tomates, etc.). Pour traiter des relations, il faudra également des ensembles de n -uplets d'individus tels que ces n -uplets se ressemblent sous un certain aspect; par exemple on aura un ensemble de couples se ressemblant en ceci que le premier membre du couple aime le second membre par exemple les couples <Quasimodo, Esmeralda>, <Romeo, Juliette>, <Juliette, Roméo>, etc.

On pourra définir le cadre nominaliste modéré comme une structure $\langle \mathcal{R}, \mathcal{I} \rangle$ où \mathcal{I} est un ensemble d'urelements – appelons-les les individus, et où \mathcal{R} est un ensemble $\{\mathcal{R}_1, \dots, \mathcal{R}_n, \dots\}$ où chaque \mathcal{R}_n est un ensemble d'ensembles de n -uplets d'individus, et tel que chaque n -uplet d'individus apparaisse dans au moins un membre de \mathcal{R}_n . Intuitivement, les membres de \mathcal{R}_n sont des ensembles de n -uplets d'individus, tels que ces n -uplets se ressemblent les uns les autres sous un certain aspect. La condition finale selon laquelle chaque n -uplet d'individus doit apparaître dans au moins un membre de \mathcal{R}_n exprime le fait que chaque n -uplet d'individus se ressemble au moins à lui-même.

J'appellerai *ensemble basique* d'un cadre un ensemble dont la construction ne peut pas être entièrement déterminée par les autres éléments du cadre. Cette notion sera utile pour appréhender ce qui est *fondamental* dans un cadre donné. Dans le cadre nominaliste modéré, \mathcal{I} n'est pas un ensemble basique : en effet, si l'on vous donne l'ensemble \mathcal{R} , vous pouvez construire \mathcal{I} . (Il suffit par exemple de construire \mathcal{I} comme l'union de tous les membres de \mathcal{R}_1 .) À l'inverse, \mathcal{R} est bien un ensemble basique : si l'on ne vous donne que \mathcal{I} , vous ne pouvez pas construire \mathcal{R} .

Ce qui est fondamental, donc, ce n'est pas l'ensemble des individus, mais la structure de ressemblance que présentent ces individus. Toutefois, les individus sont bien les entités du modèle puisque ce sont les urelements. Les nominalismes fort et modéré s'accordent donc sur les urelements (les individus), mais pas sur les ensembles basiques : \mathcal{I} est basique pour le nominalisme fort et pas pour le

modéré. Cela revient à dire que les nominalismes fort et modéré s'accordent sur les entités (les blocs de construction de la réalité), mais sont en désaccord quant à la structure fondamentale de la réalité : pour le nominalisme fort, il n'y a aucune structure, tandis que pour le nominalisme modéré, il y a une structure de ressemblance (exprimée par l'ensemble \mathcal{R} dans le cadre que j'ai présenté).

Les modèles d'une sémantique nominaliste modérée doivent avoir la forme suivante : $\langle \mathcal{R}, \mathcal{F}, \delta \rangle$. Pour interpréter les langages du premier ordre, cela ne change pas grand-chose par rapport à la sémantique nominaliste forte (c'est-à-dire la sémantique standard) : δ associe les constantes individuelles à des membres de \mathcal{F} et les prédicats n -adiques à des ensembles de n -uplets de membres de \mathcal{F} , et les règles sémantiques seront les mêmes que précédemment. Mais on peut remarquer que la présence de l'ensemble \mathcal{R} dans le modèle permet d'identifier un type spécial de prédicats : la dénotation d'un prédicat n -adique peut appartenir à \mathcal{R}_n , autrement dit un prédicat peut dénoter une classe de ressemblances. On peut appeler de tels prédicats des « prédicats réels » au sens où ils opèrent un découpage de la réalité qui correspond à une structure réelle. (Une telle affirmation n'aurait pas de sens de le cadre nominaliste fort : tous les prédicats y font un découpage arbitraire de l'ensemble des individus.) Cet aspect de la sémantique offre une ressource qui devient particulièrement utile et intéressante si l'on cherche à interpréter des langages plus riches comme ceux du second ordre.

Cadre réaliste modéré

Selon le réalisme modéré, le monde est fait d'états de choses. Un état de choses est l'instanciation d'un universel n -adique par n -individus. Il y a donc des universaux et des individus, mais tout universel est instancié et tout individu instancie au moins un universel.

Le cadre réaliste modéré pourrait être défini comme la structure $\langle \mathcal{S}, \mathcal{U}, \mathcal{F} \rangle$ où \mathcal{F} est un ensemble d'urelements qu'on appellera les individus, où \mathcal{U} est un ensemble $\{\mathcal{U}_1, \dots, \mathcal{U}_n, \dots\}$ tel que chaque membre de \mathcal{U} est un ensemble d'urelements disjoint de chaque autre et de \mathcal{F} ; on appellera les membres d'un ensemble \mathcal{U}_n des universaux n -adiques. Enfin, l'ensemble \mathcal{S} est un ensemble de couples dont le premier terme est un universel n -adique (c'est-à-dire un membre de \mathcal{U}_n) et dont le second terme est un n -uplet d'individus (c'est-à-dire des membres de \mathcal{F}), et tel que chaque universel et chaque individu apparaisse au moins une fois dans un des membres de \mathcal{S} . Chaque membre de \mathcal{S} représente ainsi un état de choses, et tout individu et tout universel apparaît dans au moins un état de choses.

Cette condition (que tout individu et tout universel apparaisse dans au moins un état de choses) a pour conséquence que ni \mathcal{U} ni \mathcal{F} ne sont des ensembles basiques dans ce cadre. En effet, si \mathcal{S} est donné, on peut construire \mathcal{F} (puisque chaque individu apparaît au moins une fois dans l'un des membres de \mathcal{S}) ainsi que \mathcal{U} (chaque \mathcal{U}_n correspond à l'ensemble des premiers termes des membres de \mathcal{S} dont le second terme est un n -uplet). À l'inverse, les seuls ensembles \mathcal{F} et \mathcal{U} ne déterminent pas la construction de \mathcal{S} . Donc \mathcal{S} est bien basique.

Dans ce cadre, l'ensemble \mathcal{F} et les ensembles \mathcal{U}_n sont des ensembles d'urelements, mais ils ne sont pas basiques; seul \mathcal{S} l'est : cela traduit bien le fait que les individus et les universaux sont les entités, les blocs de construction de la réalité, mais la seule structure fondamentale de la réalité est la structure en états de choses.

Essayons maintenant de formuler une sémantique pour la logique du premier ordre à partir de ce cadre réaliste modéré. Un modèle d'une telle sémantique devra avoir la forme suivante : $\langle \mathcal{S}, \mathcal{U}, \mathcal{F}, \delta \rangle$. Comme toujours, δ associe les constantes individuelles à des membres de \mathcal{F} , mais pour la dénotation des prédicats, les choses vont être différentes par rapport à l'approche standard : δ associera les prédicats n -adiques à des membres de \mathcal{U}_n , c'est-à-dire des universaux n -adiques. En effet, si l'on adopte un point de vue réaliste, le prédicat *rouge* devrait dénoter non pas un ensemble d'individus rouges (comme dans la sémantique nominaliste), mais un universel de rougeur. Dès lors, la règle sémantique pour les formules atomiques sera différente elle aussi. La formule Pa sera vraie ssi $\langle \delta(P), \delta(a) \rangle \in \mathcal{S}$, (autrement dit, Pa est vraie ssi il y a un état de choses qui lie l'universel dénoté par P et l'individu dénoté par a).

Un réaliste modéré pourrait bien entendu préférer utiliser la sémantique standard pour la logique du premier ordre, parce qu'elle est standard et plus simple. Mais il faut bien noter que cela reviendrait à *ne pas prendre la sémantique au sérieux* : la sémantique standard pour la logique du premier ordre ne décrit pas la relation monde/langage telle que le réaliste modéré la conçoit. (De son point de vue, ce qui rend vraies les formules du langage est un monde d'états de choses, pas un monde d'individus seulement, et les prédicats dénotent des universaux dans des états de choses, pas des ensembles d'individus.) La sémantique réaliste modérée pour la logique du premier ordre est certes plus complexe, mais elle présente ceci d'intéressant qu'elle s'efforce de représenter adéquatement la relation monde/langage du point de vue du réaliste modéré. Notons en outre que cette sémantique pour les langages du premier ordre n'est pas strictement équivalente à la sémantique standard; en effet, le schéma qui suit serait un schéma de formules valides dans la sémantique réaliste modérée telle que je l'ai décrite, mais n'en est pas un dans la sémantique standard :

$$\exists x Kx \quad \text{pour tout prédicat } K$$

Et plus généralement :

$$\exists x_1 \dots \exists x_n K^n x_1 \dots x_n \quad \text{pour tout prédicat } K^n$$

(La validité de ces schémas est une conséquence du fait que tout universel est instancié.)

On peut enfin signaler que les ressources plus étendues qu'offre le cadre réaliste modéré deviennent particulièrement utiles pour interpréter les langages du second ordre, et de façon générale, les langages plus riches; les différences avec les interprétations nominalistes de ces mêmes langages deviennent alors beaucoup plus nombreuses et significatives (Giraud, 2014).

Cadre réaliste fort

Selon le réalisme fort, il y a des universaux et ceux-ci existent indépendamment des individus, donc il peut y avoir des universaux non instanciés; par contre les individus instancient forcément des universaux. Admettons que cette dernière condition signifie que chaque individu instancie au moins un universel monadique.

Un cadre réaliste fort peut être représenté comme une structure $\langle \mathcal{U}, \varepsilon, \mathcal{F} \rangle$ où \mathcal{F} est un ensemble d'urelements qui représentent les individus, où \mathcal{U} est comme

précédemment un ensemble $\{\mathcal{U}_1, \dots, \mathcal{U}_n, \dots\}$ dont les membres sont des ensembles d'urelements tous disjoints les uns des autres et de \mathcal{F} ; on appelle les membres de \mathcal{U}_n des universaux n -adiques. Enfin ε est une fonction associant certains membres des ensembles \mathcal{U}_n à un ensemble non vide de n -uplets de membres de \mathcal{F} , et telle que pour tout i membre de \mathcal{F} il y a un u membre de \mathcal{U}_1 tel que $i \in \varepsilon(u)$. Intuitivement, ε représente l'instanciation des universaux par des individus. (Notons que ε associe *certaines* universaux, non pas tous, ce qui permet qu'il y ait des universaux non instanciés, et la dernière condition exprime le fait que tout individu instancie au moins un universel monadique.)

Quels ensembles sont basiques? (On peut considérer la fonction ε comme un ensemble de couples, donc la question se pose aussi pour ε .) Étant donné \mathcal{U} et \mathcal{F} , il est impossible de déterminer ε . Par exemple, dans un cadre très simple contenant seulement les universaux *Rouge* et *Vert* et l'individu i , la donnée de ces deux universaux et de cet individu ne permet pas de déterminer si i instancie *Vert* ou *Rouge* ou les deux. Tout ce que l'on sait, c'est que i instancie au moins l'un des deux universaux (car tout individu instancie au moins un universel monadique). Ainsi, ε n'est pas constructible à partir des autres éléments du cadre; ε est donc une fonction basique, ce qui revient à dire que l'instanciation est une structure fondamentale de ce cadre.

À l'inverse, étant donné ε , il est possible de construire \mathcal{F} puisque nous avons supposé que chaque membre de \mathcal{F} instancie au moins un universel monadique : tous les membres de \mathcal{F} vont donc apparaître dans les valeurs de ε . Ainsi, \mathcal{F} n'est pas basique; l'ensemble des individus n'est pas une structure fondamentale.

Peut-on aussi construire \mathcal{U} à partir de ε ? Non. Considérons un cadre dans lequel \mathcal{F} contient seulement l'individu i et où ε associe seulement l'universel monadique *Rouge* à i . Peut-on savoir à partir de ces seules données que \mathcal{U}_1 ne contient aucun autre universel que *Rouge*? Non, car il pourrait y avoir des universaux non instanciés : peut-être *Vert* est-il un autre membre de \mathcal{U}_1 qui n'apparaît pas dans les arguments de la fonction ε . L'ensemble \mathcal{U} n'est donc pas déterminé par ε ; c'est un ensemble basique. L'ensemble des universaux est bien un ensemble fondamental.

Les cadres réalistes fort et modéré s'accordent donc sur les entités : il y a des universaux et des individus. Mais ils sont en désaccord sur les structures fondamentales : pour le modéré, ni \mathcal{F} ni \mathcal{U} n'est basique, c'est \mathcal{S} qui l'est, c'est-à-dire que l'ensemble des universaux et celui des individus ne sont pas des structures fondamentales, c'est la structure en état de choses qui l'est; tandis que pour le réalisme fort, \mathcal{F} n'est pas basique, mais \mathcal{U} l'est, ainsi que la fonction ε , c'est-à-dire que l'ensemble des universaux et la structure d'instanciation sont les structures fondamentales de la réalité.

Esquissons maintenant une sémantique réaliste forte pour la logique du premier ordre. Un modèle dans une telle sémantique devrait avoir la forme suivante : $\langle \mathcal{U}, \varepsilon, \mathcal{F}, \delta \rangle$ où δ est une fonction de dénotation associant les constantes individuelles à des membres de \mathcal{F} et les prédicats n -adiques à des membres de \mathcal{U}_n . La règle sémantique pour les formules atomiques serait la suivante : Pa est vraie ssi $\varepsilon(\delta(P))$ est définie et $\delta(a) \in \varepsilon(\delta(P))$; autrement dit, Pa est vraie ssi le prédicat P dénote un universel qui est instancié par l'individu dénoté par a .

Cette sémantique semble-t-elle trop compliquée? Certes, elle l'est davantage que la sémantique standard de la logique du premier ordre, mais un réaliste fort ne peut pas prendre la sémantique standard au sérieux : du point de vue réaliste, ce qui rend vraie Pa n'est pas le fait que a est le nom d'un individu qui appartient à un ensemble d'individus dénoté par P ; c'est plutôt le fait que P est le nom d'un universel et a le nom d'un individu tel que cet individu instancie cet universel.

REMARQUES CONCLUSIVES

Encore une fois, prendre la sémantique au sérieux revient à considérer qu'elle doit décrire la structure de la relation monde/langage. La sémantique standard pour la logique du premier ordre ne saurait donc être prise au sérieux, à moins que l'on soit nominaliste fort. Cela dit, rien n'empêche d'utiliser la sémantique standard comme un moyen simple et pratique d'interpréter les langages du premier ordre, mais il faut alors bien reconnaître qu'on utilise la sémantique comme un outil formel sans pertinence ontologique.

Définir un cadre ontologique présente l'intérêt suivant : toute sémantique construite à partir de ce cadre, pour quelque langage que ce soit, aura précisément l'engagement ontologique déterminé par ce cadre. Autrement dit, le cadre ontologique délimite exactement les ressources qu'un défenseur de cette ontologie peut et doit utiliser pour interpréter un langage quel qu'il soit. Par exemple : si je suis réaliste modéré et si je veux interpréter le langage modal, mon interprétation ne devra pas utiliser d'autres ressources que les états de choses composés par des individus et des universaux, autrement dit je dois construire mon interprétation dans le cadre $\langle \mathcal{S}, \mathcal{U}, \mathcal{F} \rangle$. Si je parviens à construire une sémantique satisfaisante pour le langage modal à partir de ce cadre, j'aurai bien produit une interprétation du langage modal en accord avec mon réalisme modéré. Essayons d'esquisser une telle interprétation à titre d'exemple.

En s'inspirant de la théorie combinatoire de la possibilité d'Armstrong (Armstrong, 1984), on pourrait procéder de la façon suivante à partir du cadre $\langle \mathcal{S}, \mathcal{U}, \mathcal{F} \rangle$: on définit l'ensemble \mathcal{S}^* à titre d'ensemble des couples $\langle X, Y \rangle$ tel que $X \in \mathcal{U}_n$ et Y est un n -uplet de membres de \mathcal{F} . Cet ensemble \mathcal{S}^* représente toutes les combinaisons possibles d'universaux n -adiques et de n -uplets d'individus, donc (dans l'optique d'Armstrong) cela représente tous les états de choses possibles. On définit ensuite l'ensemble des mondes possibles \mathcal{W} à titre d'ensemble des parties de \mathcal{S}^* moins l'ensemble vide, c'est-à-dire n'importe quelle combinaison d'états de choses. (Cette esquisse est très imparfaite sur certains points; il faudrait ajouter plusieurs contraintes pour que cela corresponde plus exactement à la théorie d'Armstrong, mais cela donne au moins une première idée de la manière dont il faut procéder.) Il est crucial de noter que \mathcal{S}^* et \mathcal{W} ont été construits seulement à partir des éléments déjà donnés dans le cadre : les individus et les universaux; la construction de ces ensembles n'ajoute donc rien au cadre. On pourra ensuite utiliser l'ensemble des mondes possibles \mathcal{W} pour interpréter les opérateurs modaux de la façon standard. Dans la mesure où un modèle de cette sémantique aura bien toujours la forme $\langle \mathcal{S}, \mathcal{U}, \mathcal{F}, \delta \rangle$, l'engagement ontologique de cette sémantique sera en accord avec le réalisme modéré. (Les règles sémantiques pour les formules modales seront seulement beaucoup plus compliquées que pour d'autres sémantiques, mais ce n'est pas une surprise après tout : si l'on est réaliste modéré et actualiste comme Armstrong, il n'est pas surprenant que la façon dont le monde actuel rend vraies les formules modales soit assez complexe.) Et il est possible que cette logique modale réaliste modé-

rée ne soit pas équivalente à la logique modale standard ou à une autre interprétation de la logique modale construite dans un autre cadre ontologique.

Notons en outre que lorsqu'il est impossible de construire dans un certain cadre ontologique une sémantique satisfaisante pour un certain type de langage, cela constitue une objection sérieuse, soit contre une telle ontologie, soit contre l'intelligibilité d'un tel langage : il faut renoncer à l'un ou l'autre. Par exemple, la difficulté que l'on pourrait rencontrer à construire une interprétation satisfaisante de la logique modale dans un cadre nominaliste fort pourrait constituer une objection sérieuse : ou bien contre la modalité, si l'on tient à défendre le nominalisme fort, ou bien contre le nominalisme fort, si l'on tient à sauver l'idée selon laquelle les concepts modaux ont une signification.

De façon générale, les interprétations d'un même langage à partir de différents cadres peuvent différer grandement entre elles; par exemple, les quatre cadres que j'ai présentés produisent des logiques du second ordre très différentes (Giraud, 2014). Cette approche de la sémantique permet ainsi de voir comment des désaccords ontologiques se traduisent en désaccords logiques.

NOTES

¹ Voir Giraud, 2014, dont cet article est partiellement une traduction.

BIBLIOGRAPHIE

Armstrong, David M., *A Combinatorial Theory of Possibility*, Cambridge, Cambridge University Press, 1984.

Armstrong, David M., *Truth and Truthmakers*, Cambridge, Cambridge University Press, 2004.

Cameron, Ross, "Truthmakers and Ontological Commitment: Or How to Deal With Complex Objects and Mathematical Ontology Without Getting into Trouble," *Philosophical Studies*, vol. 140, no. 1, 2008, pp. 1-18.

Cameron, Ross, "How to Have a Radically Minimal Ontology," *Philosophical Studies*, vol. 151, no. 2, 2010, pp. 249-264.

Giraud, Thibaut, "Constructing Formal Semantics From an Ontological Perspective. The Case of Second-order Logics," *Synthese*, vol. 191, no. 10, 2014, pp. 2115-2145.

Heil, John, *From an Ontological Point of View*, Oxford, Oxford University Press, 2003.

Prior, Arthur, *Object of Thought*, Oxford, Clarendon Press, 1971.

Simons, Peter, "Higher-order Quantification and Ontological Commitment," *Dialectica*, vol. 51, no. 4, 1997, pp. 255-271.

Zalta, Edward N., *Intensional Logic and the Metaphysics of Intentionality*, Cambridge, A Bradford Book, the MIT Press, 1988.