OXFORD

THE ETHICS OF HUMAN ENHANCEMENT

UNDERSTANDING THE DEBATE

Edited by STEVE CLARKE, JULIAN SAVULESCU, C. A. J. COADY, ALBERTO GIUBILINI, & SAGAR SANYAL

The Ethics of Human Enhancement

The Ethics of Human Enhancement

Understanding the Debate

EDITED BY
Steve Clarke, Julian Savulescu,
C. A. J. Coady, Alberto Giubilini,
and Sagar Sanyal

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2016

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2016932104

ISBN 978–0–19–875485–5

Printed in Great Britain by
Clays Ltd, St Ives plc

Preface

We can enhance some of our mental and physical capacities above the normal upper limits for our species with the use of drug therapies and medical procedures. It is very likely that we will soon be able to enhance many more of our capacities above these limits. The prospect of enhancing human mental and physical capacities above normal upper limits has been enthusiastically embraced by several leading bioethicists and philosophers. However, a number of very prominent intellectuals, including Michael Sandel, Leon Kass, Jürgen Habermas, and Francis Fukuyama, have raised a series of concerns about human enhancement. Some of these are prudential, but others are presented as moral concerns. It has been claimed that it is morally wrong to alter human nature, that to attempt to do so amounts to 'playing God' and is a symptom of our hubris; and that we should not seek mastery of ourselves and others. It is also argued that we should heed our emotional responses to new biotechnologies, as some emotions, such as repugnance, can be a source of moral wisdom.

What are we to make of the aforementioned reactions to human enhancement? Are there well-founded moral concerns that give rise to them? Or should they simply be dismissed? Or can they be explained away—perhaps by showing that they result from a series of cognitive biases that interfere with moral reasoning? These are key questions that need to be addressed if we want to make progress in a debate that seems to have reached an impasse. By articulating reasons that can either philosophically justify or psychologically explain resistance to human enhancement, we hope to be able to open up new directions to enable debates about human enhancement to move forwards.

This volume provides a forum for philosophical reflection on debates about the ethics of human enhancement with a view to advancing those debates. The contributing authors take a careful look at resistance to human enhancement as well as sources of support for it. Some provide us with new perspectives from which to understand debates about human enhancement. Others identify points of comparison between contemporary debates about the ethics of human enhancement and other academic debates, such as that about justice for people with disabilities. Still others draw on work in psychology that informs explanations of common reactions to the prospect of enhancing humans and, more generally, altering nature and natural functions. Some of our contributors articulate original arguments either for the conclusion that we should resist human enhancement or for the conclusion that we should accept it, or at least that we should accept the use of particular enhancement technologies. Other contributors look at arguments that have been propounded either in favour of or against employing human enhancement technologies, and they seek to rebut these arguments or show that there are limits to their applicability. The collection brings

together proponents of a broad range of philosophical, and bioethical, perspectives. It includes contributions by well-known advocates of enhancement, well-known opponents of enhancement, and well-known defenders of the middle ground between outright advocacy and outright opposition to enhancement.

The volume opens with an introductory essay by Alberto Giubilini and Sagar Sanyal (Chapter 1). These authors survey the main lines of argument in recent discussions of human enhancement in bioethics and philosophy, with a particular focus on moral objections. Recent work in moral psychology, which has been invoked by some scholars to explain common reactions to the prospect of human enhancement, is also discussed. The remaining sixteen essays in the collection are organized into two sections of eight chapters each. Section I focuses on new ways of understanding debates about the ethics of human enhancement, while Section II aims to take those debates in new directions.

Section I, 'Understanding the debate', commences with three essays on the possible implications of recent work in moral psychology for human enhancement debates. This work emphasizes the importance of the intuitive and emotional components of moral judgement. It often also involves downplaying the role of deliberative reasoning in moral judgement. According to Jonathan Haidt (2012), the predominant role is to provide post hoc rationalizations for judgements that were formed earlier by emotion and intuition. Another way of approaching moral psychology, associated with Joshua Greene (2013), also gives prime position to intuition and emotion but is less dismissive of deliberative reasoning. Greene argues that while intuitive reactions are speedy, they are fallible and are only reliable in situations that our ancestors are likely to have encountered in the period of evolutionary history during which human intuitive reactions were shaped. When it comes to addressing novel moral problems, we are unable to rely on intuition and must rely on a general-purpose faculty of reason to provide moral guidance instead.

A feature of much of the debate in moral psychology is the presumed opposition between reason on the one side, and emotion and intuition on the other. This opposition is also characteristic of much bioethical debate about human enhancement, where liberals are often depicted as being on the side of reason and conservatives on the side of intuitions and emotions. In Chapter 2, C. A. J. (Tony) Coady critiques this understanding of moral judgement. He explores some of the historical discussion of the separation between reason and emotion, or between reason and the passions in philosophy, and investigates key lines of debate in contemporary discussion of the relations between emotion and reason. He criticizes work in neuroscience that invokes reason and emotion, and that proclaims putative lessons for ethics from experiments that report subjects' decisions regarding described 'moral dilemmas'. He suggests that a careful appreciation of the roles of reason and emotion in moral philosophy could improve the quality of debate about moral psychology/neuroethics and thus improve discussion of the ethics of human enhancement that are informed by moral psychology.

Chapter 3 by Joshua May explores a specific connection between the contemporary moral psychology literature and human enhancement debates. While some writers have countered 'disgust-advocates', such as Leon Kass, by arguing that disgust is morally irrelevant (to the permissibility of human enhancement), May argues that disgust is psychologically irrelevant to moral judgement formation. He scrutinizes psychological experiments on disgust and moral judgement to make two main points. One, experiments consistently link disgust to a slight change along a sliding scale of moral acceptability. They do not register anything close to a change in the valence or polarity of moral judgements (from right to wrong). Two, insofar as disgust affects moral judgement, it does so either via the normal endogenous process of moral judgement or via an abnormal exogenous mechanism. If it does so via the former, then we have reason to take seriously claims about the repugnance of human enhancement. May argues, however, that the relevant studies are better interpreted as showing that disgust acts via an external mechanism. This makes the effect of disgust on judgement akin to a performance error, as when fatigue is shown to affect moral judgements.

In Chapter 4, Doug McConnell and Jeanette Kennett examine Leon Kass' assertion that repugnance is the emotional expression of deep wisdom, and the implication that, therefore, intuitions generated by repugnance should be taken as guides to our adoption, or rejection, of enhancement technologies. Against Kass, McConnell and Kennett argue that plausible accounts of rational and wise action involve the integration of intuition and reflection. The wise only rely on intuitions over reflective thought when those intuitions have been developed through reflection, training and experience, and are subject to reflective oversight. Therefore the normative authority of intuitions is dependent on long-term reflective training. At the heart of wisdom, they argue, are the policies of epistemic humility, open-mindedness, and a willingness to justify one's actions. Repugnance alerts us to the need for caution when contemplating the introduction of enhancement technologies but it does not have the normative authority to determine our policies.

The next three essays move away from moral psychology, and consider some distinctive aspects of conservative opposition to particular forms of human enhancement. Michael Sandel (2004) relies on a therapy/enhancement distinction to ground his opposition to human enhancement and is committed to an account of a normal human, along with the explicit assumption that normal human traits are good. The disability advocacy literature has long discussed normality and its connection to the good. Linda Barclay explores a point of contact between the disability advocacy literature and the literature on the ethics of human enhancement in Chapter 5. The social model of disability holds that the extent to which traits, no matter how unusual or abnormal, limit a person's abilities and success is largely a matter of the social environment in which that person finds themself. Suppose the opponent of human enhancement adopts a nuanced understanding of normality and ability, informed by the social model of disability. Does that enable them to accept Sandel's view that parents should be 'open to the unbidden'? Barclay explores this issue, critically examining recent work by Elizabeth

Barnes. Some impairments lead to identifiable hardships when the impaired person engages in a specific social environment. Yet the impairment may not lead to an overall reduction in quality of life. For both empirical and philosophical reasons, parents face great uncertainty in assessing the effects of impairments, or of corresponding biomedical interventions, on the overall quality of life. While this means parents should heed Sandel's call to be cautious and humble in these situations, it does not require that they be open to any and all 'unbiddens'.

It is a familiar charge from opponents of enhancement that some kinds of human enhancement would amount to playing God. In Chapter 6, John Weckert explores what this charge might mean in a secular context. One possibility is to interpret it as the charge that we abandon our appropriate place in nature when we move into (non-human) nature's territory. There are two ways to unpack this interpretation: one that excludes humans from the category of nature and the other that includes humans as part of nature. Neither is satisfactory, however, as the category of the natural does no real interpretive work on either interpretation. A second possible way of understanding the charge of playing God draws on the distinction between human control and lack of control. Weckert builds up to the claim that playing God involves expanding the sphere over which we have control, to the extent that we no longer have the knowledge or capacity to competently make decisions with respect to the expanded part. Given the competence clause, the same set of actions may constitute playing God at one time but not at a later time when humans have developed sufficient competence in the relevant area.

In Chapter 7, John McMillan compares two episodes of conservative resistance to the implications of technological change: resistance to assisted insemination in the first half of the twentieth century and to biomedical enhancement today. What the two have in common, he argues, is a sort of conservatism about institutions which are bearers of value, such as the institution of marriage and the institution of essentially symmetrical relations between parent and offspring generations. He adopts Gerald Cohen's (2011) account of value conservatism as a bias in favour of retaining what is of value, even in the face of the opportunity to replace it with something of greater value. He contrasts this with a 'critical morality' which revises existing institutions, when necessary, to ensure that they conform to independent moral judgements. Thus equipped, McMillan returns to consider conservative opposition to human enhancement and finds that we can better understand it. However, conservatives are found wanting by McMillan in their account of the relevant institutions to be conserved, of the features that are essential rather than contingent to those institutions, and of why people who do not share the bias in favour of those institutions nonetheless have reason to preserve their essential features.

In Chapter 8, Chris Gyngell and Michael J. Selgelid worry that debates about human enhancement are often impeded by a lack of clarity regarding the term 'enhancement'. Identifying seven different accounts of 'enhancement' which are current in the literature, they argue that different definitions of enhancement help frame debates about the

ethics of capacity-altering biotechnologies in different ways, drawing our attention to morally relevant spectrums. They further argue that if we understand an intervention to be an enhancement to the degree that it involves improvement over a (relatively) high level of functioning, and an intervention to be a treatment to the degree that it involves improvement over a (relatively) low level of functioning, then the concept of enhancement will have moral significance. Interventions at the treatment end of the treatment-enhancement spectrum will tend to be equality promoting, and those at the enhancement end will tend to increase inequalities, all things being equal.

To round off Section I, Robert Sparrow (Chapter 9) takes up the issue of whose interests matter in a decision about whether or not to enhance the next generation. There are three contending perspectives: the parents, the child, and 'the world'. Whether an enhancement is advisable may be judged quite differently from each perspective. There are difficulties in choosing between, or balancing, the competing sets of interests. For a start, as soon as we introduce aggregate welfare (of the world) into the choice, there is the danger that this will outweigh welfare effects on an individual (the child) or on a few (the parents). A particularly worrisome possibility arises once Sparrow introduces the state as an agent. An enhancement that is reasonable from the perspective of the parent or child may be unreasonable from a societal standpoint. The likely mechanism for managing this tension is delegation of some regulatory and enforcement power to an independent party, such as the state. Sparrow notes that states tend to exercise their power in the interest not of 'the world' but of the nation. When states define the national interest, they tend to represent the nation as the political expression of the interests of 'the people', and to represent 'the people' as racially distinct from other peoples. Given these tendencies of the state, Sparrow cautions that the widespread use of human enhancement technologies may reinvigorate racism.

Section II, 'Advancing the debate', begins with an essay by Rebecca Roache and Julian Savulescu (Chapter 10). These authors advance a proposal for a type of enhancement that, they suggest, should appeal to conservatives—namely, enhancing virtues that conservatives (as opposed to liberals) regard as particularly important. Examples include human dignity and the capacity to appreciate the giftedness of life, as well as those characteristics of humanity that are considered worth preserving and which Francis Fukuyama gathers under the catch-all label of 'Factor X'. This proposal challenges conservatives to ask themselves whether enhancement would be problematic per se (and therefore problematic even in the case of the enhancement of Factor X), or problematic because of undesirable consequences—that is, because it would yield 'too much' of any enhanced trait (including, perhaps, too much Factor X). By suggesting that conservatives have reason to enhance 'conservative virtues' above current levels, Roache and Savulescu lay the foundations for a new possible dialogue between liberals and conservatives about forms of enhancement which conservatives do not seem to have good reason to oppose. Two main themes in this chapter are closely related to themes exemplified in the next two essays: the idea of designing certain desirable virtues

through enhancement, and the attempt to reconcile conservative values with at least some forms of human enhancement.

Designing virtues in our offspring that conform with our values is the focus of Bernadette Tobin's work in Chapter 11. This time the challenge is for advocates of genetic engineering in general. Tobin builds on Alasdair MacIntyre's idea of a table of socially desirable virtues that we would want our genetically engineered offspring to have, were we to design their traits through genetic interventions. Such traits include the ability to live with uncertainty, the capacity to engage in non-manipulative relations, the ability to find a vocation in one's work, the readiness to accept one's own death, the capacity to hope, and a willingness to take up arms. According to MacIntyre and Tobin, a paradox lies in the fact that those very same traits, and particularly the rejection of non-manipulative relations, would make our descendants the kind of people likely to be appalled by the idea that their ancestors had genetically manipulated them.

An attempt to reconcile conservatives' values with at least some forms of human enhancement—different from the one presented by Roache and Savulescu—is presented by Jonathan Pugh, Guy Kahane, and Julian Savulescu in Chapter 12. Here, the theoretical basis for the conservative assumption that it is inherently wrong to radically alter human nature is questioned. The authors explore a new theoretical basis for resisting alterations to human nature—one that appeals to a kind of partiality to humanity. They describe what they consider the most plausible understanding of what partiality to humanity involves, elaborating on the work of Gerry Cohen and Bernard Williams. They argue that a reasonable understanding of partiality to humanity does not rule out the use of enhancements per se. At most it sets limits to the range of enhancement that we should be willing to endorse.

An appeal to the importance of human nature when thinking about enhancement is also presented by Nicholas Agar (Chapter 13). He discusses a particularly controversial form of human enhancement—namely, the transfer of human minds, and with them personal identity, into computers. This transferring, or 'uploading', if it becomes possible, would enhance human cognitive performance by removing limitations imposed by our embodied nature, such as exposure to disease or ageing. Agar, well known for his liberal stance on enhancement, adopts a more critical attitude to mind-uploading. He argues that the latter would not be a form of human enhancement because it would fail to transfer important aspects of our personal identities (which are distinctively human) into computers.

One reason why appeals to human nature in debates about enhancement are problematic is that they can be put forward in support of opposing stances on the ethics of enhancement. While conservatives appeal to the normative status of human nature to argue against enhancement, liberals have sometimes drawn normative conclusions in favour of enhancement from considerations about the unfairness of nature. Some liberals have suggested that enhancements should be used to make up for natural inequalities imposed on us by the 'genetic lottery'. In Chapter 14, Michael Hauskeller questions the legitimacy of appeals to the unfairness of nature. He argues that it does

not make sense to declare a natural condition, which is not the result of human agency, to be 'unfair'. But if nature cannot be considered unfair, then pro-enhancement arguments based on an appeal to the unfairness of the genetic lottery cannot be sound.

The malleability of human nature through enhancement would give us the ability to reshape not only our dispositions and traits but with them our social and political structures. This ability is discussed by Steve Clarke in Chapter 15. He takes issue with Buchanan's dismissal of a conservative argument against enhancement based on risks to the stability of social structures that could arise as a consequence of altering human nature. According to proponents of some influential conservative arguments, consideration of our cognitive limitations and the complexity of our social structures should lead us to conclude that there is a significant risk that modifications of human dispositions through enhancement would backfire and produce more harms than benefits to society. Contra Buchanan, Clarke argues that conservatives are right to point out that the relationship between some of our biological dispositions and some of our social institutions—and democratic institutions in particular—is fragile. Some enhancements might create shifts in the balance of power between competing groups within society that could threaten democracy. However, Clarke argues, this fragility does not represent a reason to oppose enhancement *tout court* but rather a reason to be cautious about the adoption of some enhancements.

One form of human enhancement that is very likely to affect the political structures of our societies and, in principle, affect them for the better, is moral enhancement—that is, the use of biomedical means to improve our moral dispositions and emotions. Moral enhancement is the focus of the final two essays. In Chapter 16, Gregory E. Kaebnick suggests that moral enhancement is a very unusual type of enhancement that might appeal even to those who oppose the general idea of altering human nature through enhancing biotechnologies. Almost everyone is in favour of more moral behaviour. However, there are different ways of understanding what would constitute a moral enhancement. Kaebnick highlights how different views about the nature of morality and the notion of autonomy have different implications for whether, and how, moral enhancement should be carried out. He suggests that these different views make it difficult to know how to conduct moral enhancement and this consideration should lead us to be cautious about endorsing proposals to enhance morality.

It might well be, however, that moral enhancement is not only difficult to define and to bring about but actually not the best way to make moral progress. This is the thesis defended by Russell Powell and Allen Buchanan in Chapter 17. These authors challenge proponents of moral enhancement by defending an evolutionary account of morality which conflicts with the account often assumed by proponents of moral enhancement. Such proponents have tended to depict morality as being dominated by in-group concerns and a lack of compassion for out-groups. However, Powell and Buchanan argue that at some point in our evolutionary history our moral psychology underwent an 'inclusivist shift' that is inconsistent with that depiction. The authors do not deny that we need to make moral progress to meet the challenges we face

today in a globalized world but they suggest that these challenges are better met by cultural moral enhancement than by moral enhancement.

The work that led to this volume was supported by Australian Research Council Discovery Grant DP130103658, 'Moral Conservatism, Human Enhancement and the "Affective Revolution" in Moral Psychology', held by Steve Clarke, Julian Savulescu, and C. A. J. Coady. Alberto Giubilini and Sagar Sanyal were employed as researchers on this grant between 2013 and 2015. Most of the chapters in the collection began life as papers presented at two events organized by the five editors of this volume. The chapter by Alberto Giubilini and Sagar Sanyal began life as a background paper circulated to participants in these events. We were also fortunate enough to be able to add chapters by Michael Hauskeller and Gregory E. Kaebnick.

The two events organized by the editors were the workshop Human Enhancement: The Moral Challenge, held at the University of Melbourne on 21–22 March 2014, and the conference Human Enhancement: Destiny or Disaster?, held at the University of Melbourne on 3 July 2015. The latter event was filmed by Adam Ford from Humanity+ Australia. Footage of all the presentations has been made available on Ford's YouTube channel. Both events were supported by funding from the Australian Research Council and from the Oxford Centre for Neuroethics, which has been funded, in turn, by the Wellcome Trust's Biomedical Ethics Strategic Awards programme. Both events were organized through the Centre for Applied Philosophy and Public Ethics, which is jointly hosted by Charles Sturt University and the University of Melbourne.

As well as thanking the Australian Research Council and the Wellcome Trust for their generous support, the editorial team would like to thank Russell Blackford, Karim Jebari, Simon Longstaff, Francesca Minerva, Justin Oakley, and Suzanne Uniacke for their participation in our project events. We would also like to thank Hayley Buckley, Matthias Butler, Adam Ford, Michelle Lamb, Neil Levy, Justin Oakley, Jonathan Pugh, Rebecca Roache, Ian Sharpe, Lydia Shinoj, Elizabeth Stone, Mary Walker, and Miriam Wood for helping us, in different ways, to put together this volume. Lastly, we would like to thank Peter Momtchiloff for his expert editorial guidance.

Steve Clarke
Julian Savulescu
C. A. J. Coady
Alberto Giubilini
Sagar Sanyal

References

Cohen, G. A. (2011). 'Rescuing Conservatism: A Defense of Existing Value', in R. J. Wallace, R. Kumar, and S. Freeman (eds), *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*. Oxford: Oxford University Press, pp. 203–30.

Greene, J. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York: Penguin.

Haidt, Jonathan (2012). *The Righteous Mind*: *Why Good People are Divided by Politics and Religion*. New York: Pantheon/Knopf.

Sandel, Michael (2004). 'The Case against Perfection'. *The Atlantic Monthly*: 51–62.

Contents

Notes on Contributors

Nicholas Agar is Reader in Philosophy at Victoria University Wellington, New Zealand. His main research interests are in the ethics of the new genetics. He has also published on personal identity, environmental ethics, and the philosophy of mind. His books include *Truly Human Enhancement: A Philosophical Defense of Limits* (MIT Press, 2013) and *Humanity's End: Why We Should Reject Radical Enhancement* (MIT Press, 2010).

Linda Barclay is a Senior Lecturer in the Philosophy Department at Monash University, Australia. Her research interests lie within contemporary political philosophy, with a particular focus on discrimination and inequality. She has published on gender, nationalism, and multiculturalism, and has recently produced a number of articles on disability.

Allen Buchanan is James Duke Professor at Duke University, USA, and also holds research positions at the University of Arizona Freedom Center, USA, and King's College Law School, UK. His current research interests are in the ethics of enhancement and of synthetic biology, the philosophy of international law, and social moral epistemology. His work on enhancement includes the book *Beyond Humanity* (Oxford University Press, 2011).

Steve Clarke is Associate Professor in the Centre for Applied Philosophy and Public Ethics, Charles Sturt University, Australia, and a Senior Research Associate of the Uehiro Centre for Practical Ethics at the University of Oxford, UK. He is the author of *The Justification of Religious Violence* (Wiley-Blackwell, 2014). He is also a Co-editor of Clarke, S., Powell, R., and Savulescu. J. (eds), *Religion, Intolerance and Conflict: A Scientific and Conceptual Investigation* (Oxford University Press, 2013).

C. A. J. Coady is Emeritus Professor of Philosophy at the University of Melbourne, Australia. He is a scholar of international repute on epistemology, and on political violence and political ethics. Among his books are *Testimony: A Philosophical Study* (Oxford University Press, 1992) and *Messy Morality: The Challenge of Politics* (Oxford University Press, 2008). In 1990 he founded the Centre for Philosophy and Public Issues at the University of Melbourne, the first research centre in Australia concerned with broad issues of philosophy and public affairs. In 2005 he presented the Uehiro Lectures on Practical Ethics, and in 2012 the Leverhulme Lectures, both at the University of Oxford, UK.

Alberto Giubilini is currently a Postdoctoral Research Fellow at the Uehiro Centre for Practical Ethics and Oxford Martin School at the University of Oxford, in the UK, working on the project 'Collective Responsibility for Infectious Disease'. He has published

on various topics in bioethics (abortion, euthanasia, organ donation, conscientious objection, informed consent, family ethics) in leading journals, including *Bioethics, Journal of Medicine and Philosophy, Kennedy Institute of Ethics Journal, Journal of Medical Ethics, Hastings Center Report, Australasian Journal of Philosophy*, and *Philosophy Compass*, and has written a book (in Italian) on the value of autonomy in bioethics.

Chris Gyngell is a Marie Skłodowska-Curie Fellow with the Uehiro Centre for Practical Ethics, UK. His research interests lie primarily in bioethics, moral theory, and the philosophy of health and disease. He is currently working on a Marie Skłodowska-Curie Fellowship funded project entitled 'Selecting, Creating and Modifying Embryos', which will investigate the ethical and legal implications of new reproductive technologies, such as the gene editing technique clustered regularly-interspaced short palindromic repeats (CRISPR).

Michael Hauskeller is Professor of Philosophy at the University of Exeter, UK. He specializes in moral philosophy but has also worked in various other areas, including the philosophy of mind, the philosophy of art and beauty, phenomenology, and the philosophy of human enhancement. So far he has published 13 monographs, 6 edited volumes, and more than 80 research papers. Important publications include *Biotechnology and the Integrity of Life* (Ashgate, 2007), *Better Humans? Understanding the Enhancement Project* (Acumen, 2013), *Sex and the Posthuman Condition* (Palgrave Macmillan, 2014), and *The Palgrave Handbook of Posthumanism in Film and Television* (2015, as editor).

Gregory E. Kaebnick is Editor of *The Hastings Center Report*. He is interested in questions about the values at stake in developing and using biotechnologies, and particularly in questions about the value given to nature and human nature. He is the author of *Humans in Nature: The World as We Find It and the World as We Create It* (Oxford University Press, 2014). He is the principal investigator on the research project 'Values in Impact Assessment', funded by the National Science Foundation.

Guy Kahane is Deputy Director of the Uehiro Centre for Practical Ethics and of the Oxford Centre for Neuroethics at Oxford University, UK. He is an Associate Professor in the Faculty of Philosophy, and Fellow and Tutor in Philosophy at Pembroke College, Oxford. His research interests include value theory, meta-ethics, moral psychology, and the philosophy and ethics of neuroscience and psychology. His work on enhancement includes *Enhancing Human Capacities* (Wiley-Blackwell, 2011, co-edited with Julian Savulescu and Ruud ter Meulen).

Jeanette Kennett holds a joint appointment between the Philosophy Department and the Macquarie Centre for Cognitive Science, Australia. Among her books is *Agency and Responsibility: A Common-Sense Moral Psychology* (Clarendon Press, 2001). Her current research focus is on the contribution the sciences can make to our understanding

of the cognitive and affective underpinnings of moral reasoning, moral judgement, and moral agency, and the implications for meta-ethics and moral psychology.

Doug McConnell recently completed his PhD at Macquarie University, Australia. His research interests include moral psychology, bioethics, and applied philosophy, particularly in relation to addiction and psychiatry. His latest publication, 'Narrative Self-Constitution and Recovery from Addiction' (*American Philosophical Quarterly*, forthcoming), investigates the effect of self-conceptual content and structure on self-governance.

John McMillan is Professor and Director of the Bioethics Centre at the University of Otago, New Zealand. His books include *Empirical Ethics in Psychiatry* (Oxford University Press, 2008), co-edited with Widdershoven, Hope, and Van der Scheer; *The Limits of Consent* (Oxford University Press, 2009), co-edited with Corrigan, Liddell, Richards, and Weijer; and *Responsibility and Psychopathy: Interfacing Law, Psychiatry, and Philosophy* (Oxford University Press, 2010), co-edited with Malatesti. He is Series Editor of *Research Ethics Forum*, a book series published by Springer. He is Reviews Editor for, and an editorial board member of, *Journal of Applied Philosophy*.

Joshua May is Assistant Professor at the University of Alabama at Birmingham, USA. His research is in ethics and epistemology, particularly as these intersect with the sciences of the mind (as in neuroethics). His main project is to determine how we know right from wrong and act according to our moral judgements. His recent publications have appeared in the *Australasian Journal of Philosophy*, *Canadian Journal of Philosophy*, *Journal of Medical Ethics*, *Philosophical Studies*, and *Synthese*.

Russell Powell is Assistant Professor at Boston University, USA. He has held research fellowships at Oxford University, UK, and at Johns Hopkins University, USA. He has diverse research interests in the philosophy of biological and biomedical science, philosophical bioethics, the philosophy of medicine, legal and political philosophy, and the philosophy of technology. His publications have appeared in journals including *British Journal for the Philosophy of Science*, *Journal of Philosophy*, *Journal of Medicine and Philosophy*, *Biology and Philosophy*, *European Journal for the Philosophy of Science, American Journal of Bioethics*, *Behavioral and Brain Sciences*, *Journal of Political Philosophy,* and the *Kennedy Institute of Ethics Journal.*

Jonathan Pugh is a Postdoctoral Research Fellow in Applied Moral Philosophy on the Wellcome Trust Investigator Award in Medical Humanities project 'Neurointerventions in Crime Prevention: An Ethical Analysis'. He completed a DPhil project at Oxford University, UK, in 2014 entitled, 'Autonomy, Rationality and Contemporary Bioethics'. His research interests lie in value theory and applied ethics. He has written primarily on issues concerning personal autonomy in applied ethics, the ethics of human embryonic stem cell research, and the use of enhancement technologies.

Rebecca Roache is Lecturer in Philosophy at Royal Holloway, University of London, UK. Her research interests span ethics, metaphysics, the philosophy of mind, and the philosophy of language. Her current projects include an investigation of how enhancement technology might affect punishment, and a book about swearing. Her work has appeared in *The Philosophical Quarterly*, *Analysis*, *Philosophy*, *Behavioral and Brain Sciences*, and *Philosophical Psychology*. She is Associate Editor of *Journal of Medical Ethics*. She writes regularly for Oxford University's *Practical Ethics* blog and frequently discusses her work in the media.

Sagar Sanyal is a Research Fellow working on the Australian Research Council Project 'Moral Conservatism, Human Enhancement and the "Affective Revolution" in Moral Psychology' at the Centre for Applied Philosophy and Public Ethics at the University of Melbourne, Australia. His research interests include the ethics of enhancement, the ethics of war, and global justice. His publications have appeared in *The Journal of Philosophy*, *Philosophy Compass* and *International Journal of Applied Philosophy*.

Julian Savulescu is Professor and Director of the Uehiro Centre for Practical Ethics at Oxford University, UK. He is also Director of the Oxford Centre for Neuroethics and Sir Louis Matheson Distinguished Visiting Professor at the Centre for Human Bioethics, Monash University, Australia. His areas of research include the ethics of genetics, research ethics, new forms of reproduction, and medical ethics. In 2009 he contributed to the prestigious Tanner Lectures in Human Values, and he presented at the Sydney Festival of Dangerous Ideas. His publications include *Unfit for the Future: The Need for Moral Enhancement* (Oxford University Press, 2012), co-written with Ingmar Persson.

Michael J. Selgelid is Professor of Bioethics and Director of the Centre for Human Bioethics, and Director of the World Health Organization Collaborating Centre for Bioethics therein, at Monash University, Australia. His research focuses on ethical issues associated with biotechnology and public health ethics. He edits a book series in public health ethics analysis for Springer and another in practical ethics and public policy for ANU Press. He is Co-editor of *Monash Bioethics Review* and Associate Editor of *Journal of Medical Ethics*.

Robert Sparrow is Professor of Philosophy in the School of Philosophical, Historical and International Studies at Monash University, Australia. He has varied research interests within bioethics, political philosophy, and applied ethics. Most recently his engagement with enhancement issues has involved moral bioenhancement, eugenics, and sex selection.

Bernadette Tobin is Reader in Philosophy and Director of the Plunkett Centre for Ethics, a joint centre of Australian Catholic University and St Vincents & Mater

Health Sydney, Australia. Her research interests include the philosophy of religion, ethics, and the philosophy of education.

John Weckert is Emeritus Professor and was previously Professorial Fellow at the Centre for Applied Philosophy and Public Ethics at Charles Sturt University, Australia. He is the founding editor-in-chief of the Springer journal *Nanoethics: Ethics for Technologies that Converge at the Nanoscale*.

1

Challenging Human Enhancement

Alberto Giubilini and Sagar Sanyal

1.1 Human Enhancement: Mapping the Terrain

In bioethics the term 'human enhancement' refers to any kind of genetic, biomedical, or pharmaceutical intervention aimed at improving human dispositions, capacities, and well-being, even when there is no pathology to be treated (Giubilini and Sanyal 2015; Juengst and Moseley 2015). Selecting embryos before implantation during in vitro fertilization (IVF) procedures, inserting or deleting gene sequences, taking enhancing drugs for better physical or mental performance, pursuing life extension through stem cell applications and regenerative medicine are all examples of interventions that are, or might soon be, technically possible. The completion of the Human Genome Project at the beginning of the new millennium has provided a great deal of knowledge about human DNA and about the correlation between certain gene sequences and certain phenotypic traits. This knowledge holds promise for making genetic human enhancement not only feasible but also tailored to people's specific needs and desires.

There are two main ways of conceptualizing human enhancement with regard to the aims and scope of medicine. In general, those who oppose human enhancement think that, by going beyond what is currently seen as normal health, enhancement is beyond the proper ends of medicine (President's Council on Bioethics 2003; Pellegrino 2004; Sandel 2004). Those in favour typically endorse a welfarist approach to medicine, according to which therapy and enhancement share the same aim of increasing 'the chances of leading a good life in the relevant set of circumstances' (Savulescu, Sandberg, and Kahane 2011, p. 7), which is the aim of medicine broadly understood (Savulescu 2007; Harris 2009; Kamm 2009; Savulescu, Sandberg, and Kahane 2011). The difference between these two main conceptions of enhancement is rooted in different stances on the moral relevance of the distinction between therapy and enhancement. The non-medicine approach takes the distinction to be factually and morally important, while the welfarist approach sees the 'normality' that medicine aims to restore as a merely statistical notion, subject to change over time, with no moral significance (Bostrom and Roache 2008).

Positions regarding the morality of enhancement can be divided into three main categories (Giubilini and Sanyal 2015):

1. Some authors believe that people should be free to enhance themselves (and, according to some, also their offspring) through various means (including—should it become possible—engineering embryos). While these scholars do not endorse indiscriminate enhancement, they tend to highlight the positive aspects of enhancement and propose measures to counteract possible negative externalities. For example, some suggest that enhancements should not be imposed by the state, or that they should be subsidized for certain groups of people. Authors in this category are sometimes referred to as 'bioliberals'.
2. Some people take a more moderate approach. While they do not have any objection to enhancement in principle, they place more emphasis on the distinction between objectionable and non-objectionable forms of enhancement. Various considerations inform their cautious approach, including distributive concerns about the wealthy having access to enhancements that are not accessible to the poor, and concerns about the safety of specific techniques of enhancement.
3. Finally, some scholars have in-principle objections to human enhancement. An in-principle objection can be based either on the idea that enhancements in certain spheres of activity violate the rules that define the activity in question (e.g. doping in sport) or on the idea that human nature has a special status which enhancement would violate, particularly when carried out in radical forms, such as genetic design of offspring. Some of the language invoked at this more restrictive end bears recognizable relation to the tradition of conservative social thought, particularly terms such as 'sanctity', 'playing God', and 'human dignity'.

Some opponents of human enhancement have explicitly cast themselves as conservative. They have drawn connections to the broader conservative tradition. Therefore, in spite of some problematic implications of applying political categories to the bioethical debate,[1] we reserve the adjective 'conservative' for the third type of position, and we characterize non-conservative arguments for restrictions on enhancement (which support the second position in the taxonomy provided above) as 'restrictive'. We speak

[1] According to Ruth Macklin, the liberal/conservative dichotomy in bioethics would imply that we should consider as conservative some feminists who oppose assisted reproductive technologies (Macklin 2006, pp. 34–5). The category 'conservative' also does not overlap with the category of 'bioconservative' (which only refers to the bioethical domain). For instance, Jonathan Moreno talks of 'bioconservative' to refer to anti-enhancement positions, whether they are from the political right or political left. On the right he sees bioconservatism as the expression of a concern for the erosion of traditional values and dignity, while on the left he sees it as resulting from worries about social inequality and ecological problems (Moreno 2011, p. 121). Arthur Caplan tries to avoid political connotation and talks of meliorist and antimeliorist positions, rather than liberal and conservative positions (Caplan 2009).

of 'proponents of enhancement' or 'permissive positions regarding enhancement' instead of 'bioliberals' to avoid the impression that these positions are either entailed by the tradition of political liberal thought or are only available to those with liberal commitments.

While conservative opposition to human enhancement can be seen as a monolithic position based on a well-defined set of ethical stances (which will be reviewed in the sections below), restrictive and permissive positions are characterized by internal differences and significant internal disagreements. Such disagreements depend on different factors. For example, they might depend on the type of enhancement in question: some people who are generally in favour of enhancement oppose specific types of enhancement, such as moral enhancement (e.g. Harris 2011). Or internal differences might concern the degree of the enhancement in question: some people who are generally in favour of enhancement distinguish between more radical and impermissible forms of enhancement that might lead to a new species (post-humans) and more moderate and acceptable forms of enhancement which remain within the range of normality for our species (e.g. Agar 2013). Or, again, some might say that certain methods of enhancement are more problematic than others—for example, that genetic manipulation of embryos is more problematic than selecting a certain embryo in IVF procedures, since only in the first case would the intervention change the genetic identity of an identifiable individual.

In this introductory chapter we will focus mainly on conservative positions and, to a lesser extent, on restrictive positions. We will discuss permissive positions only to illustrate how they might respond to some typical conservative or restrictive objections. Section 1.2 provides a very brief historical overview of the conservative tradition and places the bioethics and human enhancement debates in this context. Sections 1.3–1.6 cover some recurring themes in moral conservative opposition to human enhancement. Section 1.3 looks at claims from some conservatives that bioethics debates generally, and the human enhancement debate specifically, have been restricted by assumptions that have been variously characterized as 'liberal', 'analytic', or 'instrumentalist'. It is alleged that the perspectives on enhancement recently raised by conservative writers had been excluded by the debate. Section 1.4 looks at the charge that enhancement, especially genetic enhancement, amounts to playing God. Section 1.5 looks at the moral relevance of intuition and emotion, focusing in particular on the idea of the 'wisdom of repugnance', which some conservatives have used to oppose certain uses of biotechnologies. Section 1.6 looks at appeals to human nature and to dignity. In Section 1.7 we turn to some reasons to restrict or oppose human enhancement that are not peculiar to the conservative tradition. This serves both as a way to round off the survey of opposition to enhancement, and as a contrast that helps us see what is truly distinctive about the opposition from the moral conservative camp.

1.2 Conservatism: from the French Revolution to the Biotechnology Revolution

Conservatism is, loosely, a philosophy of reaction to developments that threaten tradition (see e.g. Hirschman 1991; Scruton 2001). It stemmed from the aristocratic opposition to the egalitarian ideas of the French Revolution. According to the eminent historian of conservatism Russell Kirk,

> the essence of social conservatism is preservation of the ancient moral traditions of humanity. Conservatives respect the wisdom of their ancestors […]; they are dubious of wholesale alteration. They think society is a spiritual reality, possessing an eternal life but a delicate constitution; it cannot be scrapped and recast as if it were a machine. (Kirk 1994, p. 8)

Today, these kinds of concern are often raised with respect to one of the key priorities of conservatism: bioethics (Levin 2003, p. 53). Addressing the challenges posed by bioethics has required conservatives to focus more specifically on moral issues rather than on social or political issues. The meaning and value of life and death, the notion of personhood, the extent to which human life can be used as a commodity or a means for someone's ends—all are core concerns of what can be labeled moral conservatism in bioethics.

More recently, scientific advances, particularly in the field of biotechnologies, have provoked a new set of ethical discussions among moral conservatives in relation to human enhancement.

With the possibility of using biotechnologies to alter natural human capacities comes a set of concerns not previously expressed in bioethical debates. As Charles Krauthammer put it, rather than concerns about using and destroying early forms of human life, '[w]hat really ought to give us pause about research […] is what monsters we will soon be capable of creating' (Krauthammer 2002, p. 202). With human enhancement, the issues at stake concern the moral status of human nature, the meaning of human dignity, the permissibility of playing God, and modifying what some people see as a gift (from God or from nature) (Kass 1997; Fukuyama 2002; Levin 2003; Sandel 2004, 2007; Cohen 2006). Also, moral conservatives have questioned the permissibility or wisdom of discussing potential applications of biotechnology. Some fear that such a discussion might break certain taboos that it might be better to leave untouched (Levin 2003).

Contemporary conservatives in bioethics share a 'moral anthropology' that Eric Cohen (2006) has condensed into the following points: the recognition of a special dignity in all human beings, which is an essential feature of human nature rather than something based on contingent properties (e.g. rationality or self-awareness); a recognition of the special meaning of our mortality as the sign of the mystery surrounding human experience—that is, 'something beyond the living person's comprehension, something we cannot finally master' (Cohen 2006, p. 49); a set of principles and values about the nature of the family, involving precise prescriptions about the modalities of

procreation which describe as immoral 'various technological possibilities—like human cloning, gamete engineering, and the creation of man-animal hybrids—that would exert novel parental control over the genetic makeup of new life' (Cohen 2006, p. 50); the importance of human flourishing based on authentic human experience and efforts (as opposed to artificially enhanced minds and bodies); and an agenda for public policy that proposes, among other things, bans on creating human embryos for research purposes, non-traditional and 'unnatural' ways of creating babies, the creation of human-animal chimera embryos, and buying and selling human body parts.

Whereas Cohen lists substantial aspects, there are formal aspects that are at least as distinctive of the moral conservative movement. Ruth Macklin (2006), a critic of the conservative movement in bioethics, lists the following features as characteristic of what she calls 'the new conservative movement in bioethics' (Macklin 2006, p. 34): a propensity to use poetic and metaphoric language; repeated appeals to emotion, sentiment, and intuition; what she calls a 'mean spirited rhetoric'; and a tendency to see a proliferation of projects—that is, of concerted efforts by organized groups which conservatives oppose (e.g. the 'eugenic project', the 'biotechnology project', and so on).

1.3 Instrumentalist versus Substantive Bioethics

Some conservatives have characterized the existing bioethics debates generally, as well as the human enhancement debate specifically, as either liberal or analytic in a sense that contrasts with their own approach to the same issues. Michael Sandel suggests that liberals reach first for the language of rights, autonomy, and fairness, but that some of the hardest questions facing bioethics and the enhancement debate are not readily elucidated in these terms (Sandel 2007, pp. 9–10). Leon Kass finds the analytic approach to bioethics wanting:

> With its capture by analytic philosophy, […] [bioethics] has by and large come to content itself with analyzing moral arguments, reacting to new technological developments and taking on emerging issues of public policy, all performed with a naïve faith that the evils we fear can all be avoided by compassion, regulation and a respect for autonomy. […] its practitioners, with few exceptions, have turned the big human questions into pretty thin gruel. (Kass 1997, p. 18)

In a sympathetic history of the American President's Council on Bioethics under Leon Kass, Adam Briggle argues that the council's lasting value was to broaden the sorts of question considered in bioethics beyond certain instrumental ones (Briggle 2010). The alternative to this instrumentalist approach is described as a 'rich' or 'substantive' bioethics that extends discussion to identifying and assessing the implicit ends of certain uses of technology.

Consider a parent's decision whether to genetically enhance an embryo. The parent may implicitly assume the end of giving the child the best shot at being smart, tall, or sociable. The liberal approach, it is charged, would focus on the parent's liberty to choose whether to enhance, and on the potential harm to the future child. A more

substantive approach, the argument runs, would reconsider the wisdom of allowing all parents to aim for the end of greater intelligence, height, or sociability. If there is pervasive use of genetic enhancement of embryos, then over time the distribution of abilities among the population changes too. In due course, it may become irresponsible for parents *not* to enhance their offspring, as the failure to do so would place their children at a disadvantage compared with the new norm for intelligence, height, or sociability (Briggle 2010, p. 66).

1.4 Playing God

Proponents of human enhancement are often accused of playing God with human nature. The charge is not necessarily based on religious worldviews. There is a metaphorical sense of the expression 'playing God' which makes the objection available to non-theists as well. Such metaphorical use of the playing God objection is based on the idea that some interventions would violate some intrinsic sanctity of nature, which does not derive from God (for a discussion, see Peters 2002).

In general, the accusation refers to an incautious attitude displayed by human enhancement advocates. Enhancing human nature while disregarding the potential and/or unknown risks reveals a certain hubris. Humans are not omniscient and benevolent, and might therefore overlook the risks of tampering with genes.

Michael Sandel famously advanced a secular version of the playing God objection to human enhancement through genetic engineering (Sandel 2007). For him, one limitation of humans' drive to mastery is a failure to appreciate what he calls the *giftedness* of human life, or, with a more emphatic phrase, its 'openness to the unbidden'. Sandel thinks that this failure would jeopardize humility and solidarity among humans because mere chance (to which we all are equally subject, according to him) would be replaced by humans' choices and hyper-agency in determining exactly what kinds of people will exist. This hyper-agency would generate an 'explosion' of responsibility that might be destabilizing for our ethics—making the individual responsible for many more features of their situation (or their offspring's situation) than they are at present.

The playing God objection can also take the form of a warning that genetic manipulation technologies would blur existing species boundaries—that we would be able to create new species (Annas, Andrews, and Isasi 2002). There are some difficulties involved in sharpening this worry. Eric Juengst notes that we cannot literally preserve the species against all genetic change because genetic profiles associated with a species do change, as existing individuals pass away and new ones are born. Quite without any exogenous intervention in the process, the typical genome of a given species is likely to vary both over time and across populations that are geographically separated with little interaction. If we want to hold one species-typical genome as sacrosanct, then there is an arbitrary decision to take a snapshot at a particular point in time and space to yield the definition of species-typical (Juengst 2009, p. 50). This arbitrariness seems to make the sense of sacrosanct inviolability less plausible.

For some, the worry is not about species integrity per se but about what the biological features of a species mean for its moral status (Annas, Andrews, and Isasi 2002). For instance, it might be thought that human rights attach to individual humans because of their possessing some species-typical characteristics. A subspecies of humans created by genetic intervention (or even by enhancement more generally) might come to possess the relevant characteristics to such a heightened degree that it no longer makes sense to assign ordinary humans as much moral status as the new subspecies. For a critique of this worry, see Douglas (2013).

1.5 The Wisdom of Intuitions and Emotions

1.5.1 Intuition, Emotion, and Mystery

Some conservatives place great weight on intuitions and emotions such as disgust, repugnance, and revulsion. They claim that through disgust, a functioning moral agent recognizes moral violations when seeing the unthinkable happen (Kekes 1998, p. 106). Leon Kass famously defended intuitions and emotions as offering reliable moral guidance in the field of biotechnologies. He suggested that there is wisdom in our feelings of repugnance towards certain biotechnologies, such as human cloning and IVF (Kass 1997). Our inability to express what exactly is wrong with incest or cannibalism does not mean that these practices are acceptable, he urges. Similarly, the inability to fully articulate our opposition to biotechnologies does not constitute sufficient reason to endorse things such as human enhancement, medically assisted reproduction, or human cloning. Kass acknowledges that 'revulsion is not an argument' and grants that 'some of yesterday's repugnancies are today calmly accepted'. Nonetheless, he adds that in crucial cases 'repugnance is the emotional expression of deep wisdom, beyond reason's power to fully articulate it' (Kass 1997, p. 20). 'To pollutions and perversion,' he continues, 'the fitting response can only be horror and revulsion; and conversely, generalized horror and revulsion are *prima facie* evidence of foulness and violation' (Kass 1997, p. 21).

Another example of reliance on intuitions and emotions, rather than rational arguments,[2] is the conservative approach to the idea of human dignity. The role of dignity in the enhancement debate will be discussed more extensively in the Section 1.6. Worth noting here is that moral conservatives are often very explicit in admitting that the full appreciation of human dignity is something we grasp through intuition rather than something for which we advance reasons and arguments. The fact that many of them devote tens and tens of pages (President's Council on Bioethics 2008, particularly Kass' contribution at pp. 297–331) to discussing something that they think cannot be rationally demonstrated can be seen as either surprising or, conversely, consistent with

[2] However, as C. A. J. Coady argues in Chapter 2, it might be wrong to think of reason and emotions as forming a dichotomy. Because some emotions do have cognitive content, they are essential to practical rationality.

their assumption that no simple, straightforward definition of dignity can be provided. Thus, for instance, Eric Cohen candidly admits that, 'of course, we can never prove rationally that all human beings possess equal dignity or that human beings possess any dignity at all' (Cohen 2006, p. 48). Meilaender finds the only secular grounds for the idea of dignity in the 'mystery of human condition' (Meilaender 2008, p. 265); the same idea of mystery seems to lurk behind Francis Fukuyama's claim that what grants humans their special dignity is some not-better-clarified factor X—that is, 'some essential human quality underneath' which is what remains 'when we strip all of a person's contingent and accidental characteristics away' (Fukuyama 2002, p. 149). In his discussion of human dignity, Leon Kass claims that, despite the difficulties in clearly articulating reasons in support of the idea of dignity, 'we can in fact recognize dignity, both when we see it shining and when we see it extinguished' (Kass 2008, p. 306).

Emotions and intuitions also ground moral sanctions of what is perceived as polluting holiness and corrupting the venerable, in the form either of God or of the natural order. Such sanctions are often connoted as *taboos* (Levin 2003). Many conservatives take taboos to be important for the functioning and stability of societies because of the 'inarticulate awe' they inspire in many people (Levin 2003, p. 64). Because taboos are often grounded in strong intuitions or emotions, there is a tension between the desire to preserve them and the desire to engage in public debates on a taboo topic. Preserving a taboo implies refraining from rational analysis of it, even when the analysis aims to defend the core of the taboo. Yet many conservatives want to engage in public debate, which requires arguments and publicly accessible evidence and reasons. Yuval Levin has called this 'the paradox of conservative bioethics':

> Bioethics is necessarily focused on the deepest and most sensitive of human moral intuitions and taboos […]. At the same time, it is also directed toward policy, which in a liberal democracy rightly means that it must be an ethics of fully public arguments. It is therefore in the business of public arguments about taboos. […]. Herein lies the paradox of conservative bioethics. (Levin 2003, p. 54)

Admittedly, grounding their views in intuitions and feelings of revulsion poses some problems to moral conservatives when they want to defend their positions in the public square if interlocutors do not share the intuitions and feelings—as is likely to be the case with regard to repugnance at human cloning. Indeed, it makes the conservative position less likely to succeed in influencing policy-making in liberal Western societies, where public policies are mainly based on the no-harm principle (Roache and Clarke 2009, p. 7).

Levin suggests that because the public square requires participants in public debate to provide arguments in support of a certain position, conservatives should try to provide more and better arguments. However, a closer look at his proposal seems to suggest that this seeming openness to arguments and reasons is a limited one. He claims that a conservative argument defending the truth in taboos 'should begin from a sense of what is humanly important' (Levin 2003, p. 64), which, without elaboration of how to arrive at this sense seems to bring us back to the intuitive and emotional sphere.

Aside from the difficulties related to bringing the conservative view into the public sphere, the reliance on intuitions and gut feelings defended by conservatives is problematic for at least three other reasons. First, intuitions often change over time. Initial strong opposition to IVF has mostly faded away, for instance. So an independent criterion would be required to determine when intuitions are trustworthy and when they are not. Second, society and technology have changed significantly since the time of our ancestors: our set of intuitions might have evolved to meet the demands of a past environment, and may be unsuited to today's society (Singer 2005, pp. 347–8). Third, intuitions are subject to cognitive biases that render them unreliable. This is especially true when technologies put intuitions to work under completely novel circumstances, as in the case of human enhancement (Roache and Clarke 2009). Indeed, one of the most common accusations raised against conservative positions in the enhancement debate is that of being subject to cognitive biases, such as a 'status quo' bias, and that these biases are due to unanalysed intuitions and emotions (Bostrom and Ord 2006; Kahane and Savulescu 2013; Giubilini 2015).

1.5.1.1 APPEALS TO LITERATURE

Appeals to literary scenarios or expressions are a common argumentative device among moral conservatives. An approach that is based on intuitions, emotions, and the notion of 'mystery' is more likely to find resources in poetic and rhetorical language than in argumentation and cool reasoning. Thus Fukuyama's criticism of the biotechnology revolution starts off with the mention of Huxley's *Brave New World* and Orwell's *1984* in the first five lines of the first chapter. It is not so obvious, Fukuyama says, what is wrong in Huxley's *Brave New World* in which no one is hurt and everybody has their desires immediately satisfied. Still, we believe that such a world is undesirable even if we struggle to formulate the reasons for this view (Fukuyama 2002, pp. 5–6). The implicit suggestion is that prospects opened by biotechnologies are making our world so similar to the 'brave new' one that the fact that we cannot fully articulate the reasons against it does not mean we should not oppose it.

In a similar fashion, Kass writes that 'the people dehumanized à la *Brave New World* are not miserable, don't know that they are dehumanized, and, what is worse, would not care if they knew. They are, indeed, happy slaves with a slavish happiness' (Kass 1985, p. 35, quoted in Fukuyama 2002, p. 6).

Similarly, in discussing the risk that human life might become the product of other people's choice and thus lose its gifted character, Michael Sandel describes the scenario of the movie *Gattaca*, 'in which parents routinely screen embryos for sex, height, immunity to disease and even IQ'. 'There is something troubling', he writes, 'about the *Gattaca* scenario, but it is not easy to identify what exactly is wrong with screening embryos to choose the sex of our children' (Sandel 2004, p. 53). It is easier to point out that we do not want to live in a world like that.

It is interesting to note here that Sandel's point against the 'drive to mastery' is supported not by a direct argument but by an attempt to explain why people perceive

Gattaca-like scenarios as wrong. In the case of performance-enhancing biotechnologies, for instance, he writes that

> as the role of enhancements increases, our admiration for the achievement fades—or, rather, our admiration shifts from the player to his pharmacist. This suggests that our moral response to enhancement is a response to the diminished agency of the person whose achievement is enhanced.

He therefore concludes that the idea of the giftedness of life is the only way 'to account for what we admire about human activity', and presumably what we do not admire about human enhancement (Sandel 2004, p. 54).

These examples suggest that the function of literary examples for conservative authors differs from the typical function of thought experiments in philosophical arguments. While both aim to prompt certain emotional and intuitive reactions in the audience through fictional cases, traditional thought experiments typically rely on analogies between improbable scenarios and more realistic ones to test certain theories (e.g. the doctrine of double effect or the moral difference between doing and allowing in trolley problem scenarios). Literary examples in the enhancement debate, on the other hand, are not presented as unrealistic test scenarios but as likely consequences of the application of the principles informing the pro-enhancement positions.

1.6 Human Dignity and Human Nature in the Enhancement Debate

Human dignity is frequently invoked by conservatives to argue against human enhancement. A representative overview of the conservative appeal to dignity is provided by the collection of essays *Human Dignity and Bioethics*, commissioned in 2008 by the President's Council on Bioethics. It gathers contributions by prominent contemporary conservative thinkers (as well as a few non-conservatives).

Providing a single definition of human dignity would be impossible because, as many conservatives themselves recognize, 'the term itself is abstract and highly ambiguous' (Kass 2008, p. 306; for a similar assessment, see also Fukuyama 2002, p. 148). As a result, they propose different and sometimes conflicting interpretations. Since a working definition is needed, however, we present here the general definition proposed by Patrick Lee and Robert George because it is sufficiently broad to be considered as acceptable by almost anyone:

> The dignity of a *person* is that whereby a person excels other beings, especially other animals, and merits respect and consideration from other persons. (Lee and George 2008, p. 410)

Different interpretations of human dignity depend not only on different possible definitions of 'person', 'excellence', and 'respect' but also on the different understandings of human nature and of the foundations of dignity.

1.6.1 Dignity of Human Nature

The discussion of human dignity in conservative circles assumes there is something special about human nature that confers on any human a moral status that is a) superior to that of any other living being and b) the same for all human beings (regardless of age, level of development, capacities). In the following passage, for instance, Lee and George explain 'the nature and basis of human dignity' by a direct appeal to human nature:

> The criterion for moral worth must be the possession of a property that does not itself vary in degree—it must, that is, be the possession of a *nature*. Being of moral worth must be grounded in an entity's existence as a substance of a certain sort […] rather than in the possession of a set of accidental or variable properties. (Lee and George 2008, p. 415)

In similar vein, Eric Cohen writes:

> All members of the human family—all living human bodies—have a human life, and *therefore* deserve the respect that such membership commands. (Cohen 2006, p. 47, emphasis added)

And here is Francis Fukuyama:

> The most significant threat posed by contemporary biotechnology is the possibility that it will alter human nature and thereby move us into a 'posthuman' stage of history. This is important […] because human nature exists, is a meaningful concept, and has provided a stable continuity to our experience as a species. It is, conjointly with religion, what defines our most basic values. (Fukuyama 2002, p. 7)

Liberals may object to this normative conception of human nature on various grounds. They might point out that alterations to human nature have always characterized our species (Lewens 2012); or that human nature cannot have normative value unproblematically because it contains both good and bad aspects; or that our conception of the good is independent of, and indeed is used to evaluate, human nature (Buchanan 2008).

There are, however, other ways in which human nature can be deployed to support the idea of human dignity. Appeals to human nature often ground concerns about the threat posed by enhancement to the idea of a natural equality among human beings. By sharing the same human nature, humans, qua humans, have equal dignity. Such appeals can be drawn out in either a religious or a secular vision.

As Meilaender notes, 'we are equal to each other, whatever our distinctions in excellence of various sorts, precisely because none of us is the "maker" of another one of us. We have all received our life—equally—as a gift from the Creator' (Meilaender 2008, p. 264). All that is valuable and dignified in humans is given them by God through the nature He creates, hence the normative status of human nature. If we strip away the idea of its religious aspect, the result resembles Sandel's secular 'ethics of giftedness' based on the 'appreciation of the gifted character of human powers and achievements' (Sandel 2004, p. 55). Technologies such as preimplantation embryo selection threaten to make some humans the authors of the talents and powers of others. The President's

Council on Bioethics report entitled *Human Cloning and Human Dignity* makes the point by contrasting *begetting* with *making*, suggesting that a child made according to parents' wishes would not have the same dignity as a begotten child (President's Council on Bioethics 2002, p. 112).

1.6.2 Foundations of Human Dignity

An important distinction within conceptions of human dignity is that between the comparative, or aristocratic, notion and the non-comparative, or egalitarian, notion (Lee and George 2008; Meilaender 2008). The former is the notion of the *full* 'dignity of being human', a notion conveying a sense of worthiness and nobility that is found not in every human being but only in those with certain excellences, virtues or capacities. The non-comparative egalitarian notion is the *basic* 'dignity of being human', which is shared by any form of human life but which cannot, according to Kass, be shown to have any ontological or theoretical grounds (Kass 2008, pp. 316–20).

Kass finds shortcomings in both conceptions when considered separately. The first is too exclusive and the second lacks a satisfactory justification. He therefore proposes a third notion of dignity, which he calls the 'dignity of being in-between' (Kass 2008, p. 324). The view casts humans as halfway between the other animals and God. Humans are god-like and have aspirations towards what is higher, and therefore a human is 'more than an animal' (Kass 2008, p. 323). However, humans are dependent on their embodied nature for everything high about human life. On Kass' approach, the aristocratic and egalitarian conceptions are reconciled through their mutual dependence. Human flourishing and excellence depend on the merely animal aspects of life which all forms of human life share. These animal aspects are, in turn, dignified by being integrated with the higher functions that are typical of human nature (Kass 2008, pp. 321–2). Although conceptually sophisticated, when it comes to practical guidance about altering human nature, this third conception seems to boil down to an endorsement of the non-comparative notion of dignity. As Kass writes in discussing the implications of the in-between notion of dignity, 'any man's very being requires that we respect his life' (Kass 2008, p. 323).

Eric Cohen argues that even though most conservatives on this issue come from a religious tradition, the idea of dignity informing their conservatism does not have a religious foundation (Cohen 2006). Nonetheless, it is difficult to see how a religious foundation can be turned into a secular one in some discussions of dignity.

Consider Meilaender, for instance:

> It may be that we cannot make good sense of an egalitarian and non-comparative understanding of human dignity, to which our civilization has in many ways been committed, if we abstract it entirely from the context of the religious beliefs that formed it. […] I doubt, in fact, that there is any way to derive a belief in the equal worth of every human being from the ordinary distinctions in merit and excellence that we all use in some spheres of life; it is grounded, rather, not in our relation to each other, but in our relation to God, from whom—to use a mathematical metaphor—we are equidistant. (Meilaender 2008, pp. 262–3)

Or Leon Kass:

> The dignity of being human [...] completes itself and stands tallest when we bow our heads and lift our hearts in recognition of powers greater than our own. The fullest dignity of the god-like animal is realized in its acknowledgement and celebration of the divine. (Kass 2008, p. 329)

A distinct basis for human dignity, probably inspired by the classical Kantian conception of dignity, is the intrinsic rationality that some people believe characterizes human nature (Lee and George 2008). In contemporary conservatives, this view does not require that the high properties that distinguish humans from other animals be actualized. The core of the view seems to be that all forms of human life share an intrinsically rational nature, which makes this view overlap with the Catholic view. In other words, 'because personhood is based on the kind of being that one is—a substantial entity whose nature is rational nature—one cannot lose one's fundamental personal dignity, as long as one exists as a human being' (Lee and George 2008, p. 410).

1.6.3 Views on Human Dignity among Proponents of Human Enhancement

Proponents of enhancement have two main strategies to contest dignity-based arguments against enhancement. One is to reject the concept of dignity—for instance, by claiming that it is too vague or that it adds nothing to bioethical discussion (Macklin 2003). Pinker goes further and suggests that it is a stupid concept that is relative, fungible, and even potentially harmful and deceptive (Pinker 2008).

The other strategy is to argue that enhancement may actually *promote* human dignity by improving those qualities and virtues that confer a special worth on human beings. Nick Bostrom defends a notion of human dignity centred on the qualities that enhanced humans might possess. 'Dignity as quality', as he calls it, is a 'kind of excellence', 'a virtue or an ideal, that can be cultivated, fostered, respected, admired, promoted, etc.' (Bostrom 2008, p. 175).

1.7 Restrictive (Non-conservative) Positions on Human Enhancement

So far we have focused on opposition to enhancement from what we have called the moral conservative camp. However, opposition may also stem from concerns about social equality or the negative consequences of enhancement.

1.7.1 Egalitarian Concerns

A major concern about equality is that human enhancement technologies might only be affordable to the world's wealthy, or in any case to some small privileged group, thus exacerbating the already marked inequalities between rich and poor (Mehlman and Botkin 1998; Glannon 2001; McKibben 2003; Mehlman 2003). There are also fears that enhancement carried out over several generations may create two separate species, one of which will have the power to dominate the other (Silver 1999).

One response to this objection is that we already accept significant inequalities due to differences in opportunity, and should not single out biotechnological enhancements for special criticism on this count. We accept that the rich can afford better education for their children, for instance. Nonetheless, the worry remains that enhancement may significantly worsen existing inequalities in a more radical way. As an instance of the general worry that great inequalities in society tend to undermine its stability and threaten democracy, it has also been argued that a society with significant inequalities due to genetic differences in particular would have the same socially undesirable effects (Mehlman 2003).

A second, more powerful, response from enhancement advocates is that enhancement, far from being opposed to equality, can be used to make up for the unfairness of the 'genetic lottery' (thus embracing a form of luck egalitarianism) by bringing the least fortunate up to a decent minimum of capacities and well-being. In other words, a strategy available to enhancement advocates is that of replacing a strict egalitarian with a sufficientarian notion of justice, and to argue that enhancement is required by this sufficientarian notion (Savulescu 2006). Thus we might have a policy whereby enhancements are subsidized for those who cannot afford them, so that everybody could have access to them, which would level the playing field (Mehlman 2009; Buchanan 2011).

1.7.2 Utilitarian Concerns

Utilitarian opposition to enhancement should also be considered. In particular, enhancement interventions might drain resources away from more useful medical research aimed at serious diseases that threaten the well-being of the global majority (Selgelid 2014). This problem already exists: only 10 per cent of medical research is used to address 90 per cent of the global burden of disease (the so-called diseases of the poor). However, this is no licence to make things even worse. As Michael Selgelid puts it, '[i]f it is safe to assume that treatment and prevention of serious disease would improve human well being to a greater extent than interventions aimed at things like greater than average height, then utility would be adversely affected in the process' (Selgelid 2014, p. 11).

One response to such utilitarian concerns about enhancement is to say that procreative liberty trumps utilitarian considerations because individual rights generally take priority over duties towards society. A second response is that there are other utilitarian considerations in favour of enhancement that should also be taken into account—namely, the cost to society (rather than to the individual) of failing to enhance individuals (Levy 2013).

1.7.3 Balancing Principles

Since human enhancement raises ethical concerns about individual liberty, equality, and utility, and since such considerations are often at odds with each other, Selgelid has argued for a shift in the enhancement debate. He proposes investigating principled

ways of balancing or trading off the different concerns. Such a 'moderate pluralist' theory would determine, for instance, 'how great the equality and/or utility costs of enhancement would need to be in order for liberty infringement to be justified' (Selgelid 2014, p. 12).

1.7.4 Autonomy and Moral Equality

Jürgen Habermas opposes certain sorts of enhancement because of concerns about the autonomy and moral equality of future generations. The practice of morality, he contends, is based on idealizing presuppositions that each person is responsible for giving ethical shape to their life, and that each enjoys equal treatment with complete reciprocity of rights and duties (Habermas 2003, p. 92). Social influences shape us and, in that sense, mitigate our responsibility. Viewing ourselves as autonomous and responsible requires reference to a point beyond social influences. That reference point is the moment of birth, which demarcates the person's 'natural fate' from their 'socialized fate' (Habermas 2003, pp. 59–60). Yet if even this 'natural fate' is influenced by humans, we should enquire whether this changes our understanding of autonomy and responsibility.

Viewing ourselves as equals in the shared practices of moral community, Habermas claims, is incompatible with sorts of dependence which are one-way rather than reciprocal. Having one's genome 'programmed', as he puts it, places the younger generation at the receiving end of such a one-way relation. This is fine if the younger generation can be assumed to consent to the programming, something that Habermas thinks is true of therapeutic interventions (Habermas 2003, p. 43). He contrasts genetic programming to the socializing influence of parents in child-rearing. Leaning on his wider philosophical theory, Habermas says that in the latter case, the child is involved in a reciprocal communicative process in which they can respond to the parents, which Habermas suggests does not happen in the case of the genetic enhancement of offspring. Granted, there are some permissible cases of enhancement, he thinks, such as aiming for a stronger immune system or a longer lifespan. These are highly general goals, and they can be assumed to have the consent of the future child despite the fact that the child is at the receiving end of a one-way and non-reciprocal act (Habermas 2003, pp. 51–2). For more specific enhancements, consent cannot be assumed. Consider the more general contrast with the transfer of culture down the generations. Whereas a future generation can question and selectively accept or reject elements of received culture, Habermas suggests that it is not in a position to reject the cumulative effect on the gene pool of generations of decisions guided by the forces of profit and individual preference (Habermas 2003, p. 72).

1.7.5 Unintended Bad Consequences and Futility

Some moral conservatives oppose enhancements without appealing to values peculiar to the conservative tradition. Alongside interpretations of concepts such as sanctity, dignity, and human nature, conservatives may also argue against enhancement on the grounds of overlooked bad consequences.

Consider first the so-called 'perversity thesis', whereby certain actions or policies have unintended consequences that produce the opposite of what they were meant to promote (Hirschman 1991). An example of this is the view that there is a 'precisely balanced' human nature such that any human attempt to alter it is likely to end in disaster (President's Council on Bioethics 2002, p. 287). The perversity thesis in this sense can take two different forms: one factual and the other value oriented. Francis Fukuyama argues that, because the interactions between single genes and phenotypic traits is very complex, with any gene coding for multiple proteins, altering any single gene or genetic sequence to obtain a desirable trait might have bad unintended consequences for the expression of other desirable traits (Fukuyama 2002, pp. 74–5, 92–3). Whereas this concern is based on factual claims answerable to evidence, unintended bad consequences can also be presented in more abstract terms familiar to the conservative tradition. According to Leon Kass, for instance, with biotechnologies '[w]e may get more easily what we asked for only to realize it is vastly less than what we really wanted. Worse, we may get exactly what we ask for and *fail* to recognize what it cost us *in coin of our humanity*' (Kass 2008, p. 303).

One way to address the perversity thesis is to note that it seems to involve a misconception of evolution. Consider, for instance, the idea that 'the human body and mind, highly complex and delicately balanced as the result of eons of gradual and exacting evolution, are almost certainly at risk from any ill-considered attempt at "improvement"' (President's Council on Bioethics 2002, p. 287). Powell and Buchanan argue that the force of this point rests on a misunderstanding of how evolution works (Buchanan 2011; Powell and Buchanan 2011). Evolution is not, as the quote implies, a 'Master Engineer' that makes its creation (nature) 'a *stable, completed masterpiece* that can only be ruined by any human attempt to improve it' (Buchanan 2011, p. 156). Rather 'the fact that natural selection has operated on a trait does not ensure that the trait is optimal' (Buchanan 2011, p. 156). So there is room for beneficial and ameliorating human intervention, unless opponents of enhancement can show that the current traits or dispositions are optimal (Bostrom and Ord 2006; Kahane and Savulescu 2013).

Apart from the perversity thesis, Hirschman identifies two other common argumentative strategies typical of 'the rhetoric of reaction' (Hirschman 1991). These are the 'futility thesis' and the 'jeopardy thesis', and they are often used to argue against human enhancement.

The futility thesis targets enhancement that seeks to change human nature, rather than individual human beings. Francis Fukuyama notes that it would be extremely difficult, if not impossible, to alter human nature through genetic intervention. The human gene pool is so large that such changes would have to occur to a very large proportion of individuals and on a wide range of genes, something deemed very unlikely to happen. As he explains, 'a handful of rich people genetically modifying their children for greater height or intelligence would have no effect on species-typical height or IQ.' (Fukuyama 2002, p. 79). While this claim seems to contradict what Fukuyama himself has said about the possible unintended bad consequences of altering gene sequences, it certainly is one of the argumentative strategies available to the opponents of

human enhancement who are not committed to the 'perversity thesis'. By appealing to the futility of seeking to change human nature in this way, it might be argued that enhancement for this end is a poor use of resources better spent elsewhere.

The jeopardy thesis involves a slippery slope argument. Leon Kass argues that permitting therapeutic cloning (i.e. the creation of embryos via nuclear transfer from somatic cells in order to obtain embryonic stem cells with the same DNA as the somatic cells) would open the path to reproductive cloning (Kass 1997). He writes:

> Only a few years ago, slippery slope arguments were used to oppose artificial insemination and in vitro fertilization using unrelated sperm donors. Principles used to justify these practices, it was said, will be used to justify more artificial and more eugenic practices, including cloning. Not so, the defenders retorted, since we can make the necessary distinctions. And now, without even a gesture at making the necessary distinctions, the continuity of the practice is held by itself to be justificatory. (Kass 1997, p. 24)

1.7.6 The Old Eugenics and the New

A common criticism, if not an argument per se, from opponents of human enhancement invokes comparisons with state-led eugenics programmes in the first half of the twentieth century. This has prompted distinctions between the old and the new eugenics, or between old eugenics and 'liberal eugenics' (Agar 2004).

The term 'eugenic' refers to practices aimed at promoting desirable traits in a given population (e.g. by encouraging the 'best' individuals to reproduce), or at eradicating undesirable traits (e.g. by reducing the reproduction of 'defective' individuals). When Francis Galton first coined the term 'eugenics' in 1883 (Kevles 1985, p. ix), it did not have bad connotations. It acquired negative connotations in the twentieth century due to coercive policies adopted in the US and in Scandinavian countries at the beginning of that century (including marriage restrictions of the 'unfit' and their forced sterilization), and in Germany during the Nazi era (including the killing of 'defective' children, institutionalized deficients, homosexuals, Jews, and other groups). Arguments in favour of human enhancement are often presented by moral conservatives as dangerously close to the ideas that led to Nazi eugenic programmes and other, comparable, horrors.

Proponents of enhancement distinguish it from the 'old eugenics' programmes enforced in the first half of the twentieth century. The old eugenics involved state-directed programmes based on scientifically or morally flawed premises (e.g. the alleged superiority of particular races or populations over others) and conducted by ethically unacceptable means. Contemporary proponents of human enhancement defend a 'liberal eugenics' in which people would be left free to decide whether and how to enhance themselves or their offspring (Savulescu 2001, p. 425). Some opponents of enhancement are swayed by these arguments (Fukuyama 2002, pp. 85–8) but most are not (Sandel 2004, pp. 119–20; O'Mathúna 2006; Kass 2008, p. 301; Pellegrino 2008, p. 515).

Robert Sparrow, in particular, argues that despite the good intentions of proponents, the outcomes of the new eugenics might turn out uncomfortably similar to those of the old one. Stressing the incompatibility of strong consequentialism with liberalism, he suggests that a consequentialist argument for human enhancement cannot but entail the endorsement of a unique model of human being that has the greatest chances to have a good life in a given environment. In Sparrow's words, 'in any given environment at least, there is a "best" genome, which parents are obligated to provide for their children' (Sparrow 2011, p. 36). In Western countries, for instance, it might be a white, blond, blue-eyed male (Sparrow 2011, p. 35). If a consequentialist argument for human enhancement is to be taken seriously, Sparrow thinks that proponents of liberal eugenics should also commit to allowing state imposition of this model on the population.

1.8 The Affective Revolution in Moral Psychology and the Conservative/Liberal Divide

1.8.1 Relevant Evidence from Moral Psychology and Neuroscience

Recent empirical evidence from the fields of moral psychology and neuroscience provides some reason to think that the neural and psychological correlates of conservatives' approach to certain moral issues may differ from those correlates in liberals approaching the same issues. Note that we have thus far avoided speaking of permissive views about human enhancement as specifically *liberal*. We wanted to avoid the impression that the liberal tradition entails a permissive stance, or that a permissive stance is only compatible with a liberal philosophy. The moral psychology literature discussed in this section, however, commonly uses the conservative/liberal categories, and we adopt these terms to reflect the literature. Nonetheless, the reader should retain in the back of their mind questions about the aptness of thinking of permissive positions on enhancement in these terms.

As we have seen above, Michael Sandel (2004) ascribes to liberals a narrow focus on three ethical concerns: a) for autonomy, b) for fairness, and c) for individual rights.

Interestingly, these three ethical concerns overlap with the three 'modules' that, according to Jonathan Haidt (2012), generally characterize the psychology of liberals. He defines modules as adaptive mechanisms that generate intuitive or emotional reactions to stimuli. As he puts it,

> Modules are like little switches in the brains of all animals. They are switched on by patterns that were important for survival in a particular ecological niche, and when they detect that pattern, they send out a signal that (eventually) changes the animal's behavior in a way that is usually adaptive. (Haidt 2012, p. 123)

In humans, such signals generated by modules in our brain are our moral intuitions and emotions. Haidt argues that in making moral judgements, liberals rely on modules

sensitive to issues about liberty (vs. oppression), fairness (vs. cheating), and care (vs. harm), and therefore their moral judgements are influenced by such emotions as reactance (for violations of liberty), anger (for violations of fairness), and compassion (for violations of care). These three do not exhaust the range of moral modules and emotions, however. Expression of the full range is more consistently observed in conservatives (Graham, Haidt, and Nosek 2009; Haidt 2012). The modules that activate our sensitivity towards issues concerning loyalty/betrayal (e.g. breaking promises), authority/subversion (e.g. disobedience to moral or political authorities), and sanctity/degradation (e.g. violation of the natural order or of purity) play a more prominent role in the formation of moral and political judgements of conservatives.

According to Haidt and many of his colleagues (e.g. Haidt and Joseph 2007; Greene 2008), moral judgements are often the product of intuitions or emotions. Reasoning is usually a post hoc exercise when justifying one's views to others. Many recent studies in moral psychology (Hauser et al. 2007), as well as fMRI (functional magnetic resonance imaging) investigations (Greene et al. 2001), seem to support the emotivist/intuitionist account of the formation of moral judgements. For instance, evidence suggests that some moral principles (e.g. that of double effect) that, if valid, might be used to justify one's moral judgements are not usually available to explicit moral reasoning even after a judgement has been made that seems to accord with use of the principles; some other moral principles apparently become available only in the form of post hoc rationalization for a certain judgement (e.g. the direct vs. indirect harm principle) (Cushman, Young, and Hauser 2003). These findings seem to indicate that the process by which we form a moral judgement is independent from reasoning about principles and values, and instead depends on intuitions for which the subject eventually looks for a rational justification.

Other empirical evidence has been interpreted as suggesting that feelings of disgust heavily affect, if not determine, people's moral judgements (Wheatley and Haidt 2005). For instance, people tend to make harsher judgements when in presence of foul air (Haidt 2012, pp. 70–1); and subjects who wash their hands before the experiment become more moralistic about issues related to purity. Zhong and Liljenquist (2006) call this the 'Macbeth effect'.

If this interpretation of the psychological evidence is correct, it might provide some grounds for understanding moral conservatives' positions in the enhancement debate, as presented in the sections above. For instance, the emphasis on the value of human nature and on human dignity might be explained, at least in part, by the prominent role played by the sanctity/degradation module in conservatives' formation of moral judgements. The prominent role of disgust in moral evaluation could tell us something about the role of the feeling of repugnance to which some conservatives appeal, particularly in light of studies suggesting that conservatives have greater disgust sensitivity than liberals (Rozin et al. 1999; Inbar, Pizarro, and Bloom 2009).

1.8.2 Possible Objections

There are different ways to criticize these findings on the role of intuitions and emotions in the formation of moral judgements, as well as their alleged explanatory role for conservative stances. One is to question their relevance for *moral* psychology. Another is to question the conclusiveness of the empirical evidence. A third is, more specifically, to question the claim that they actually explain the conservative stances.

As for the first type of criticism, in discussing the results of Haidt's empirical research, Jeanette Kennett and Cordelia Fine (2009) have questioned their relevance for the issue of what explains our *moral* judgements as opposed to other kinds of judgements. They note that the way we conceive of morality implies that a moral judgement, as opposed to other kinds of judgement, is such only if we have first reasoned about our intuitions and incorporated them into principles for which we can be held accountable. The conception of morality that Kennett and Fine propose is consistent with the idea that if exercising control over immediate reactions is not possible, then 'our moral concepts will lack application' (Kennett and Fine 2009, p. 85).

Josh May (2014) has questioned the results and methodology of many of the experiments that seem to show that disgust influences moral judgements. In particular, he suggests that such experiments show that disgust makes moral judgements harsher but not that it alters the polarity of judgements. In other words, the evidence currently available does not show that disgust determines people's beliefs about the permissibility or impermissibility of an action but only that it affects the perceived degree of goodness or badness, rightness or wrongness. It may be that disgust is a consequence, rather than a cause, of moral judgements (Giubilini 2016). This type of criticism is consistent with the more general idea, put forward, for instance, by Cordelia Fine (2006), that conscious reasoning and reflection might play a much more relevant role in the formation of our moral judgements than Haidt and colleagues grant (see also Payne 2005). Some work in moral psychology and neuroscience seems to point in this direction. For example, there are suggestions from fMRI studies that liberals display more brain activity in the anterior cingulate cortex—the area of the brain activated when a certain situation requires responses that diverge from one's habitual response tendency (Amodio, Jost, and Yee 2007). These data suggest that liberals have more neurocognitive sensitivity for altering habitual response patterns, so that liberal moral stances might be the result of the overcoming of immediate reactions.

The third criticism—that is, that such results from moral psychology and neuroscience cannot and should not be used to explain conservative stances—has more to do with how conservatives might want to defend themselves from an implication that seems to lurk behind these empirical studies. The implication is that conservatives are simply less rational than liberals. Thus, for instance, the *New Atlantis*, the voice of the new American conservative wave (its editors include Yuval Levin, Roger Scruton, and Eric Cohen),[3] does not hesitate to denounce as 'stupid' (*New Atlantis* 2003, p. 104) an

[3] The editors are listed at <http://www.thenewatlantis.com/about/masthead>.

authoritative study, published in the *Psychological Bulletin* of the American Psychological Association, which provides evidence for the claim that the resistance to change of conservatives is mainly due to their attempt to manage uncertainty and threat, and that conservatives are in general less tolerant of ambiguity, more dogmatic, less open to experience, and less tolerant of uncertainty (Jost et al. 2003).[4]

References

Agar, Nicholas (2004). *Liberal Eugenics: In Defence of Human Enhancement*. Oxford: Blackwell.

Agar, Nicholas (2013). *Truly Human Enhancement: A Philosophical Defense of Limits*. Cambridge, MA: MIT Press.

Amodio, D. M., Jost, J. T., and Yee, C. M. (2007). 'Neurocognitive Correlates of Liberalism and Conservatism'. *Nature Neuroscience* 10(10): 1246–7.

Annas, G. J., Andrews, L. B., and Isasi, R. M. (2002). 'Protecting the Endangered Human: Toward an International Treaty Prohibiting Cloning and Inheritable Alterations', *American Journal of Law and Medicine* 28: 151–78.

Bostrom, Nick (2008). 'Dignity and Enhancement', in the President's Council on Bioethics (ed.), *Human Dignity and Bioethics*. Washington DC: US Government Printing Office, pp. 173–206.

Bostrom, Nick and Ord, Toby (2006). 'The Reversal Test: Eliminating Status Quo Bias in Applied Ethics', *Ethics* 116(4): 656–79.

Bostrom, Nick and Roache, Rebecca (2008). 'Ethical Issues in Human Enhancement', in Jesper Ryberg, Thomas Petersen, and Clark Wolf(eds), *New Waves in Applied Ethics*. Basingstoke: Palgrave Macmillan, pp. 120–52.

Briggle, Adam (2010). *A Rich Bioethics*. Notre Dame, IN: University of Notre Dame Press.

Buchanan, Allen (2008). 'Human Nature and Enhancement', *Bioethics* 23(3): 141–50.

Buchanan, Allen (2011). *Beyond Humanity?* New York: Oxford University Press.

Caplan, Arthur (2009). 'Good, Better, or Best?', in J. Savulescu and N. Bostrom (eds), *Human Enhancement*. Oxford: Oxford University Press, pp. 199–210.

Cohen, Eric (2006). 'Conservative Bioethics and the Search for Wisdom', *Hastings Center Report* 36(1): 44–56.

Cushman, F., Young, L., and Hauser, M. (2003). 'The Role of Conscious Reasoning and Intuitions in Moral Judgment', *Psychological Science* 17(12): 1082–9.

Douglas, Tom (2013). 'Human Enhancement and Supra-personal Moral Status', *Philosophical Studies* 162(3): 473–97.

Fine, Cordelia (2006). 'Is the Emotional Dog Wagging its Rational Tail, or Chasing It?' *Philosophical Explorations* 9(1): 83–98.

Fukuyama, Francis (2002). *Our Posthuman Future. Consequences of the Biotechnology Revolution*. New York: Farrar, Straus and Giroux.

Giubilini, A. (2015). 'Don't Mind the Gap: Intuitions, Emotions, and Reasons in the Enhancement Debate', *Hastings Center Report*, 45(5): 39–47.

[4] We would like to thank Steve Clarke, Neil Levy, Jonathan Pugh, and Rebecca Roache for useful comments and suggestions on previous drafts of this chapter.

Giubilini, A. (2016). 'What in the World is Moral Disgust?', *Australasian Journal of Philosophy*, 94(2): 227–42.

Giubilini, A. and Sanyal, S. (2015). 'The Ethics of Human Enhancement', *Philosophy Compass* 10(4): 233–43.

Glannon, Walter (2001). *Genes and Future People*. Cambridge, MA: Westview Press.

Graham, J., Haidt, J., and Nosek, B. A. (2009). 'Liberals and Conservatives Rely on Different Sets of Moral Foundations', *Journal of Personality and Social Psychology*, 96(5): 1029–46.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. (2001). 'An fMRI Investigation of Emotional Engagement in Moral Judgment', *Science* 293: 2105–8.

Greene, Joshua (2008). 'The Secret Joke of Kant's Soul', in W. Sinnott-Armstrong (ed.), *The Neuroscience of Morality: Emotions, Brain Disorders, and Development*. Cambridge, MA; London: MIT Press, pp. 35–79.

Habermas, Jürgen (2003). *The Future of Human Nature*. Cambridge: Polity Press.

Haidt, J. and Joseph, C. (2007). 'The Moral Mind: How Five Sets of Innate Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps even Modules', in P. Carruthers, S. Laurence, and S. Stich (eds), *The Innate Mind: Foundations and the Future*, Vol. 3. New York: Oxford University Press: pp. 367–92.

Haidt, Jonathan (2012). *The Righteous Mind*. New York: Pantheon.

Harris, John (2009). 'Enhancements are a Moral Obligation', in J. Savulescu and N. Bostrom (eds), *Human Enhancement*. Oxford: Oxford University Press, pp. 131–54.

Harris, John (2011). 'Moral Enhancement and Freedom', *Bioethics*, 25(2): 102–11.

Hauser, M., Cushman, F., Young, L., Kang-Xin Jin, R., and Mikhail, J. (2007). 'A Dissociation between Moral Judgment and Justifications', *Mind and Language* 22(1): 1–21.

Hirschman, Albert (1991). *The Rhetoric of Reaction*. Cambridge, MA: Belknap Press-Harvard University.

Inbar, Y., Pizarro, D. A., and Bloom, P. (2009). 'Conservatives are more Easily Disgusted than Liberals', *Cognition and Emotions* 23(4): 714–25.

Juengst, E. and Moseley, D. (2015). 'Human Enhancement', in Edward N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, at <http://plato.stanford.edu/archives/sum2015/entries/enhancement>, accessed 16 June 2015.

Juengst, Eric (2009). 'What's Taxonomy Got to Do with It? "Species Integrity", Human Rights, and Science Policy', in J. Savulescu and N. Bostrom (eds), *Human Enhancement*. Oxford: Oxford University Press, pp. 43–58.

Jost, T., Glaser, J., Kruglanski, A. W., and Sulloway, F. J. (2003). 'Political Conservatism as Motivated Social Cognition', *Psychological Bulletin* 129(3): 339–75.

Kahane, Guy (2011). 'Mastery Without Mystery: Why there is No Promethean Sin in Enhancement', *Journal of Applied Philosophy* 28(4): 355–68.

Kahane, Guy and Savulescu, J. (2013). 'Normal Human Variation: Refocusing the Enhancement Debate', *Bioethics*, online first, doi:10.1111/bioe.12045.

Kamm, Francis (2009). 'What Is and Is Not Wrong with Enhancement?', in Savulescu and Bostrom (eds), *Human Enhancement*. Oxford: Oxford University Press, pp. 91–130.

Kass, Leon (1985). *Toward a More Natural Science: Biology and Human Affairs*. New York: Free Press.

Kass, Leon (1997). 'The Wisdom of Repugnance: Why We should Ban the Cloning of Humans', *The New Republic* 2 June: 17–26.

Kass, Leon (2008). 'Defending Human Dignity', in President's Council on Bioethics (ed.), *Human Dignity and Bioethics*. Washington DC: US Government Printing Office, pp. 297–332.

Kennett, J. and Fine, C. (2009). 'Will the Real Moral Judgment Please Stand Up? The Implications of Social Intuitionist Model for Meta-ethics and Moral Psychology', *Ethical Theory and Moral Practice* 12(1): 77–96.

Kekes, John (1998). *A Case for Conservatism*. Ithaca, NY: Cornell University Press.

Kevles, D. J. (1985). *In the Name of Eugenics*. Los Angeles: University of California Press.

Kirk, Russell (1994). *The Conservative Mind*. 7th rev. edn. Washington DC: Regnery Publishing.

Krauthammer, Charles (2002). 'Why Pro-Lifers are Missing the Point', in Kristol, Willam and Cohen, Eric (eds), *The Future is Now: America Confronts the New Genetics*. Lanham MD: Rowman and Littlefield, pp. 201–3.

Lee, P. and George, R. (2008). 'The Nature and Basis of Human Dignity', in the President's Council on Bioethics (ed.), *Human Dignity and Bioethics*. Washington DC: US Government Printing Office, pp. 409–33.

Levin, Yuval (2003). 'The Paradox of Conservative Bioethics', *New Atlantis* Spring: 53–65.

Levy, Neil (2013). 'There May Be Costs to Failing to Enhance, as Well as to Enhancing', *American Journal of Bioethics* 13(7): 38–9.

Lewens, Tim (2012). 'Human Nature: The Very Idea', *Philosophy and Technology* 25(4): 459–74.

McKibben, Bill (2003). *Enough: Staying Human in an Engineered Age*. New York: Henry Holt Company.

Macklin, Ruth (2003). 'Dignity is a Useless Concept', *British Medical Journal* 327: 1419.

Macklin, Ruth (2006). 'The New Conservatives in Bioethics: Who Are They and What Do They Seek?', *Hastings Center Report* 36(1): 34–43.

May, Josh (2014). 'Does Disgust Influence Moral Judgment?', *Australasian Journal of Philosophy* 92(1): 125–41.

Mehlman, M. J. (2003). *Wondergenes: Genetic Enhancement and the Future of Society*. Bloomington, IN: Indiana University Press.

Mehlman, M. J. (2009). *The Price of Perfection: Individualism and Society in the Era of Biomedical Enhancement*. Baltimore: Johns Hopkins University Press.

Mehlman, M. J. and Botkin, J. (1998). *Access to the Genome: The Challenge to Equality*. Washington DC: Georgetown University Press.

Meilaender, Gilbert (2008). 'Human Dignity: Exploring and Explicating the Council's Vision', in the President's Council on Bioethics (ed.), *Human Dignity and Bioethics*. Washington DC: US Government Printing Office, pp. 253–77.

Moreno, Jonathan (2011). *The Body Politic: The Battle over Science in America*. New York: Bellevue Literary Press.

New Atlantis (2003). 'Out of the Right Mind', *New Atlantis* 3(Fall): 103–5.

O'Mathúna, Donal (2006). 'Human Dignity in the Nazi Era: Implications for Contemporary Bioethics', *BMC Medical Ethics* 7: 2.

Payne, B. K. (2005). 'Conceptualizing Control in Social Cognition: How Executive Functioning Modulates the Expression of Automatic Stereotyping', *Journal of Personality and Social Psychology* 89(4): 488–503.

Pellegrino, Edmund (2004). *Biotechnology, Human Enhancement, and the Ends of Medicine*. The Centre for Bioethics and Human Dignity, 30 November. <http://cbhd.org/content/biotechnology-human-enhancement-and-ends-medicine>.

Pellegrino, Edmund (2008). 'The Lived Experience of Human Dignity', in the President's Council on Bioethics (ed.), *Human Dignity and Bioethics*. Washington DC: US Government Printing Office, pp. 513–39.

Peters, Ted (2002). *Playing God? Genetic Determinism and Human Freedom*. New York: Routledge.

Pinker, Steven (2008). 'The Stupidity of Dignity', *The New Republic* 8 May 2008. <http://www.newrepublic.com/article/the-stupidity-dignity>, accessed 11 November 2015.

Powell, R. and Buchanan, A. (2011). 'Breaking Evolution's Chains: The Promise of Enhancement by Design', in Savulescu, Julian, ter Meulen, Ruud, and Kahane, Guy (eds). *Enhancing Human Capacities*. Oxford: Wiley-Blackwell, pp. 49–67.

President's Council on Bioethics (ed.) (2002). *Human Cloning and Human Dignity: An Ethical Inquiry*. Washington DC: US Government Printing Office.

President's Council on Bioethics (ed.) (2003). *Beyond Therapy: Biotechnology and the Pursuit of Happiness*. Washington DC: US Government Printing Office.

President's Council on Bioethics (ed.) (2008). *Human Dignity and Bioethics*. Washington DC: US Government Printing Office.

Roache, R. and Clarke, S. (2009). 'Bioconservatism, Bioliberalism, and the Wisdom of Repugnance', *Monash Bioethics Review* 28(1): 1–21.

Rozin, P., Lowery, L., Imada, S., and Haidt, J. (1999). 'The CAD Triad Hypothesis: A Mapping between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Codes (Community, Autonomy, Divinity)', *Journal of Personality and Social Psychology* 76(4): 574–86.

Sandel, M. (2004). 'The Case against Perfection', *The Atlantic* 293(3): 51–62.

Sandel, M. (2007). *The Case Against Perfection*. Cambridge, MA: Belknap Press.

Savulescu, J. (2001). 'Procreative Beneficence: Why We Should Select the Best Children', *Bioethics* 15(5/6): 413–26.

Savulescu, J. (2006). 'Justice, Fairness, and Enhancement', *Annals of the New York Academy of Sciences*, 1093: 321–38.

Savulescu, J. (2007). 'Genetic Interventions and the Ethics of Enhancement of Human Beings', In Bonnie Steinbock (ed.), *The Oxford Handbook of Bioethics*. New York: Oxford University Press, pp. 516–35.

Scruton, Roger (2001). *The Meaning of Conservatism*, 3rd edn. Basingstoke: Palgrave.

Selgelid, Michael (2014). 'Moderate Eugenics and Human Enhancement', *Medicine, Health Care and Philosophy* 17(1): 3–12.

Silver, L. M. (1999). *Remaking Eden: Cloning, Genetic Engineering and the Future of Human Kind*. London: Phoeniz.

Singer, Peter (2005). 'Ethics and Intuitions. *The Journal of Ethics* 9: 331–452.

Sparrow, Robert (2011). 'A Not-so-New Eugenics: Harris and Savulescu on Human Enhancement', *Hastings Center Report* 41(1): 32–42.

Wheatley, Thalia and Haidt, Jonathan (2005). 'Hypnotic Disgust Makes Moral Judgments More Severe', *Psychological Science* 16(10): 780–4.

Zhong, Chen-Bo and Liljenquist, Katie (2006). 'Washing Away Your Sins: Threatened Morality and Physical Cleansing', *Science* 313(5792): 1451–2.

SECTION I

Understanding the Debate

2

Reason, Emotion, and Morality

Some Cautions for the Enhancement Project

C. A. J. Coady

> The heart has its reasons which reason knows nothing of [. . .] We know the truth not only by the reason, but by the heart. (Blaise Pascal)
>
> there is a subtlety of perception in real attachment even when it is borne towards man by one of the lower animals, which leaves the highest intellect behind. To this mind of the heart, if I may call it so, in Mr. Dick, some bright ray of the truth shot straight. (Charles Dickens, *David Copperfield*, speaking of the way in which the simple-minded Mr Dick discerns aspects of an elusive troubled personal relationship)

2.1 Introduction

It is a standard reproach in popular speech and academic contexts alike to admonish someone as being emotional or using emotive speech. There is an implied, and often enough explicit, contrast with the speaker's own reliance on reason or rationality. The speaker is cool and collected where you (the accused) are heated. The contrast, and its evaluative emphasis, has struck me in examining the 'human enhancement' debate since it is sometimes claimed that one of the divergences between so-called 'liberal' proponents of human enhancement and the 'conservative' opponents of the project is that the conservatives, or at least that group of them whose objections are, roughly speaking, inherent rather than prudential, rely on emotion whereas their liberal opponents rely on reason.

This claim often explicitly or implicitly takes it that the two human faculties are strongly opposed and that the appeal to emotion is somehow illicit. The role of emotion is often equated with 'gut feeling' and disparaged as, at best, a kind of shorthand reaction produced by evolutionary processes to allow quick responses when reason hasn't time to do the real work. This is common in the writings of neuroscientists who address themselves to ethics, and is most prominent in the work of Joshua Greene and

his colleagues (e.g. Greene 2013). It then seems to reinforce views of philosophers who want to object to 'conservative' positions in a range of areas, and notably in the enhancement debate. What I want to do here is discuss the relation between emotion and reason, thereby summarizing and joining a debate amongst philosophers that has had too little impact on the neuroscientists and their philosophical admirers. I argue that the common derogatory opposition is oversimplified and dubiously helpful. I hope this will contribute to an understanding of the nature of reasons and reasoning since, as Pascal and Dickens claim, in their different ways, a person's success in seeking truth is as much guided by the positive aspects of their emotional make-up as by the positive aspects of what we think of as their intellective life.

2.2 Sketching the Tradition

The tradition in Western thought (and elsewhere too) to the effect that reason and emotion are two distinct and even warring elements in our conscious make-up is ancient and persistent. Reason is pictured as pure, cool, higher, authoritative, and ideally regulative, whereas emotion is seen as tainted, turbulent, lower, and in need of control. Some of this opposition is linked to a mind–body dualism, though the link is not inevitable since it depends on what is included within 'mind' or 'body'; some of it arises from philosophical traditions, some of it from religious traditions.

On the philosophical side, one might cite Plato's image in the *Phaedrus* of the charioteer of reason ruling the two galloping horses of the passions. Admittedly, Plato has one of the horses (the white one) represent good emotions even though they must still be under the control of emotionless reason. Nor is this image irrelevant to Plato's developed theory of moral fulfilment as rational contemplation of the eternal (bloodless) 'Form of the Good', nor his picture of healthy civil society as the domination of unruly passionate plebs by the philosophically rational Guardians. Certain aspects of Stoic thought also contribute to the picture since they describe most of what we would call emotions (plus some other affective states) as involving inherently false judgements that must be corrected by reason. The Stoics did somewhat grudgingly recognize that there were a few 'good feelings' but their description of these is decidedly cool, dispassionate, and remote from normal human experience. For Diogenes, the condemnation of pleasure, fear, and desire is supposedly mitigated by an endorsement of 'joy' as 'a well-reasoned elation', of 'watchfulness' or 'caution' as 'a well-reasoned shrinking', and 'wishing' as 'a well-reasoned wishing' (Long and Sedley 1987).

On the religious side, the strand of Christianity that opposes 'the flesh' to the 'spirit' sometimes includes rather a lot of human emotional orientations within the category of the 'flesh' that need to be subdued by the spirit. Paul's Letter to the Galatians with its dictum that 'The desires of the flesh are against the Spirit, and the desires of the Spirit

are against the flesh' (Galatians 5:17) has been influential here. Much modern use of the term 'puritanism' refers in part to the emphasis on stern control of the unruly passions by a directive rationality (though one informed by God's Revelation).[1] It is also significant and has been stressed by many feminist philosophers, amongst whom most influentially has been Genevieve Lloyd, that there is a gendered history of the use of reason to mark off men from women and accord the former a superior position because women are dominated by emotions in a way that is not characteristic of 'the Man of Reason' (Lloyd 1984).

We might for convenience (but with an awareness of the looseness involved) call this reason tradition Rationalism, and, of course it contrasts with other intellectual traditions, notably that of Romanticism. I have mentioned the feminist critique of the masculinizing of reason, and this sometimes leads such theorists to stress the significance of emotions in the hierarchy of human characteristics. In the area of ethics, Martha Nussbaum has stressed the positive (and negative) importance of emotions for reason (Nussbaum 2001), and this is connected with the relatively recent feminist advocacy of an ethics of caring as a counterweight to rationalistic ethics. In much of this literature, the emotions underpinning parental, and especially maternal, concern are discussed. Prominent in this advocacy have been Virginia Held and Sara Ruddick, but there are many other feminist philosophers, and some male philosophers have also joined the fray, notably Michael Slote (Ruddick 1989; Held 2005; Slote 2007). Although the term 'caring' is recent as a technical tag for this theoretical concern, the histories of both philosophy and theology have seen similar emphases. One notable reaction against an excessive rationalism was that of John Stuart Mill who recorded movingly, and sometimes painfully, in his *Autobiography* the revolution in his outlook brought about by his recognition of the crucial place of emotion, poetry, nature, and the arts in a fulfilling human life. As Mill put it:

> I had now learnt by experience that the passive susceptibilities needed to be cultivated as well as the active capacities, and required to be nourished and enriched as well as guided… The maintenance of a due balance among the faculties, now seemed to me of primary importance. The cultivation of the feelings became one of the cardinal points in my ethical and philosophical creed. (Mill 1957 [1873])[2]

More recently there have been further significant moves in philosophy towards a more sympathetic and nuanced picture of the role of emotions in a morally fulfilling life. Typically this has produced a great deal of focused, distinction-laden, and often technical disputation, but the general tendency is clearly more positive than the Rationalist tradition sketched above.[3]

[1] To be fair to the original seventeenth-century Puritans, the modern usage departs a good deal from the beliefs and practices that earned those Protestant dissenters the title.

[2] My thanks to John Howes for reminding me of Mill's conversion.

[3] For a very good introduction to this debate, see Deonna and Teroni 2008.

2.3 The Turn to Neuroscience

A new development beyond philosophy has cast a different sort of light on the reason/emotion divide. This is the development within the discipline of cognitive science or neuroscience of an account of the brain's functions that shows, or purports to show, various things about the relations of emotion and reason in thinking, and especially in moral thinking and deciding. A significant moment in this development was Antonio Damasio's book *Descartes' Error: Emotion, Reason and the Human Brain*, which began in its introduction by stating:

> I had been advised early in life that sound decisions came from a cool head, that emotions and reasons did not mix any more than oil and water. I had grown up accustomed to thinking that the mechanisms of reason existed in a separate province of the mind, where emotion should not be allowed to intrude, and when I thought of the brain behind that mind, I envisioned separate neural systems for reason and emotion. (Damasio 1995, p. xi)

But this picture was destroyed for Damasio when he studied the history of Phineas P. Gage, whose life had been transformed in 1848 by a horrible accident when an explosion drove an iron rod, 43 inches long and 1.25 inches thick, weighing over 13 pounds, straight through his brain and out again, landing a good 100 feet away covered in blood and brains. Contrary to every expectation, Gage was not killed. Indeed, shortly after the accident he was capable of describing the accident and its aftermath calmly in great detail to a doctor who attended him and found him 'perfectly rational' (Damasio 1995, p. 6). But Gage's subsequent history showed that although his capacity for logical reasoning was unimpaired, his character and the emotional resources that had made him a highly respected and capable human being were largely destroyed. His reasoning about what to do, how to respond to others, and how to shape his life was dramatically impaired. He could not sustain work, became dissolute and unreliable, exhibited himself in a freak circus, and died at the age of 38. Damasio and his colleagues managed to find out a surprising amount about the type of damage to Gage's brain and the likely source of the impairments that dogged the rest of his life. Taking hints from that, Damasio conducted numerous contemporary investigations of brain damage that confirmed his dawning conviction that 'reason may not be as pure as most of us think it is or wish it were and that emotions and feelings may not be intruders in the bastion of reason at all: they may be enmeshed in its networks for worse and for better' (Damasio 1995, p. xii).

This is not the place for a résumé of Damasio's empirical studies and conclusions, impressive and stimulating as they are.[4] I cite his work merely to illustrate an interesting shift in the traditional separation of reason and emotion that I began by outlining. This shift has come from empirical investigation into behaviour and its connection with brain activity, and, although Damasio's detailed suggestions and theories are certainly controversial, the picture of an ideal reason aloof in its nature from reliance on emotion and other human affects has been, or should have been, rendered doubtful. However,

[4] Damasio (2007) has more recently discussed the relevance of neuroscience to ethics.

in recent years, there has been an upsurge in neuroscientific studies that rely on fMRI imaging and purport to show that changes in discrete areas of the brain said to be concerned respectively with emotion and reason ('cognition') influence our thinking and deciding about moral and practical matters. Some of these studies reflect a return to the Rationalist tradition described earlier since the decisions arising from the reason zone are taken to be clearly superior to the mistaken responses arising from the emotion zone. I think that most of these studies have failed to learn properly from Damasio and rely upon a picture of separate and conflicting brain areas for emotion and cognition that merely mirrors the philosophical tradition I outlined earlier. There is thus good reason to challenge the model of neuroscience at work here on its own grounds, but I shall not make that central here.[5] Many of these studies proceed by showing the strong influence that emotions allegedly have on reactions to descriptions of challenging moral scenarios (often fantasy ones invented by philosophers) and then contrast these with 'rational' reactions. Quite often the assumption is that the rational and morally right reaction is a utilitarian one; indeed, at most a somewhat narrow version of utilitarian thinking, as we shall explore below. Barton and Ritoz, for example, state that 'decisions made on the basis of deontological principles usually lead to results that are not as good as the best that could be achieved'. It is clear from the context that this is not merely a restatement of the difference between deontological and utilitarian theories but an often unargued preference for the latter. Other theorists, such as Joshua Greene and his colleagues, declare outright that all non-utilitarian judgements in moral matters are simply errors caused by emotional reactions emanating in a specific emotion centre in the brain. Sometimes this line of thought is linked to a somewhat confusing contrast of the personal and impersonal, in ways that I cannot discuss fully here.[6]

It is likely that the drive to detect errors in the moral thinking of their subjects (rather than merely describing differences and giving some neural story about them) has been influenced by the work of such psychologists as Daniel Kahneman and Amos Tversky. They have fashioned distinguished careers by demonstrating the ways in which ordinary folk make errors of reasoning in much of ordinary life, such as the estimate of risks, or the testing of their own beliefs—errors which are said to be due to intuitive and emotional reactions. I don't mean to endorse all of Kahneman and

[5] For an excellent discussion of such flaws, see Glannon 2011, especially ch. 4.

[6] Greene seems to have been the originator of this terminology for an explanation of the contrasting reactions to such trolley problem examples as bystander/driver and footbridge, the idea being that the impersonal behaviour of throwing a switch contrasts with the personal behaviour of pushing. But imagining oneself as the agent who pushes the switch on the trolley to divert it onto the line that kills one, rather than the line that kills five, or shoves the innocent person over the bridge to stop or divert the trolley, are both cases of imagining something personally done by oneself. Moreover, the divergence of reactions is likely to be much the same if one imagines the examples as done by someone else and so not personal in the first sense. And if the contrast is rather one of bodily contact versus non-bodily contact then, to cite an example of Francis Kamm's, those who object to shoving the bystander should not object to throwing a switch which springs a trapdoor that plunges the bystander on to the track, but this reaction is seriously improbable for a deontologist.

Tversky's conclusions, but at least for some of their error discoveries they have a non-contentious, mathematically robust background account in probability theory to make good the claim of error.[7] In ethics, this is largely missing, but the calculative nature of utilitarianism may contribute to the unconscious and sometimes conscious privileging of it in the quest to expose the moral errors of the vulgar mob and its faulty reliance on emotion.

However, a study by Bartels and Pizarro counters this trend, first by simply pointing out its prevalence and question-begging nature, and then by seeking to show that psychopathic personalities, who are commonly regarded as largely lacking a concern for morality and as being defective in certain emotional capacities, such as empathy, are conspicuously more comfortable with utilitarian reasoning than with deontological. Since it is also commonly held that psychopaths have plenty of rationality in the narrow rationalist sense, and since it seems to many theorists and non-theorists that morality is an activity much concerned with good and bad reasons for acting, then the fact that psychopaths are bad exemplars of moral agents means that the rationalist approach, and perhaps Utilitarianism more generally, is not a good one for understanding morality or what is really rational about it in the way of reason-giving (Bartels and Pizarro 2011).

The Bartels–Pizarro study arrived at its conclusions by testing subjects against a scale for psychopathy and then testing them for answers to a battery of 'moral dilemmas' mostly drawn or adapted from philosophers' examples. It also checked the responses against scales for what it called Machiavellianism, and 'no meaning to life' attitudes, and these also showed greater tendencies to support the utilitarian solutions to the dilemmas.

Gratifying as these conclusions are to those of us who are neither psychopaths nor utilitarians, there are reasons to be somewhat sceptical about the ambitious extent of the conclusions of the experiment, even if it is not as crude conceptually as the studies it criticizes. I shall leave aside for the moment problems regarding the nature of psychopathy about which there is considerable debate within psychiatry and psychology, and mention several other conceptual crudities in the analysis, which Bartels and Pizarro have inherited from those they disagree with.

First there is the understanding of utilitarianism which is construed in narrow act utilitarian terms, without any appreciation of the more complex versions of the theory, such as rule and other indirect forms of utilitarianism; nor is there any attempt to explore or mention a distinction between utilitarianism as a theoretical justification of morality by contrast with a decision procedure for particular actions. To be fair, many of the studies they criticize do equate moral rationality with primitive act utilitarianism.[8] This

[7] Their work has appeared in numerous academic journals, but some crucial claims are made in Tversky and Kahneman 1981.

[8] Some have objected even more strongly that the 'utilitarian' answers regarding trolley-type scenarios do not reflect any sort of utilitarian outlook. See, for example, Kehane et al. 2015. The authors point out that calling these replies 'utilitarian' ignores the utilitarian theorists' commitment to 'the greater good of all',

simplification is important because more complex utilitarian theories might well yield the same result as the 'deontological' reaction. Stephen Nathanson, for instance, has argued for an absolute moral rejection of terrorist acts on rule utilitarian grounds, even though it is usually thought that utilitarians reject all absolute moral prohibitions and allow that some terrorist acts (e.g. 'ticking bomb' situations) may be morally right (Nathanson 2010, see also Coady 2012).[9]

Second, there is no appreciation of the fact that some of the moral puzzles provided are such that many non-utilitarian theorists would give the very same answers that the so-called utilitarians give. This is a feature not only of Bartels–Pizarro but of many of the so-called 'moral dilemmas' that I have seen. The general point is that not all deontologists are absolutists about moral prohibitions and many allow, following W. D. Ross, that duties can clash and be resolved by balancing (Ross 2002.), so that the intentional killing of one person (or a few) might be permissible to save many. Elsewhere I have called these 'threshold deontologists'. Michael Walzer's category of 'dirty hands' is also hostile to utilitarianism but allows consequences and duties to override deep moral constraints only in 'supreme emergencies', thus distinguishing itself from both utilitarianism and threshold deontology. These points suggest that psychologists studying moral thinking often fall into the vulgar error of mistakenly construing any decisive concern for consequences in morality as utilitarian, whereas deontologists (and other non-utilitarians) are certainly free to consider consequences in a range of cases since they can affect what duties are in play and how they are to be weighed. So in the 'moral dilemma' studies such as the Soldiers example (leaving an injured soldier to be captured and killed when saving him would mean that all of the troop—including him—would be killed), the outcome can be viewed as presenting a clash between significant deontological duties to the soldier and to the troop, and likewise in the Driver/Bystander example (the original option in the Trolley problem: should the driver turn the trolley so that one dies rather than five?). Regarding the Driver example, non-utilitarians such as Frances Kamm give deontological reasons (to do with a Principle of Permissible Harm) for supporting the supposedly utilitarian solution, and she gives extremely complex deontological reasons for rejecting the utilitarian response in the Footbridge example, where an innocent passerby is thrown over a bridge to death in front of the trolley in order to save the five (Kamm 2007, pp. 23–32,

which leads utilitarians to argue for highly altruistic and demanding moral self-sacrifice in other contexts. This certainly contrasts utilitarians with psychopaths, who characteristically are indifferent or hostile to altruistic demands of any sort. But these points do not show that a narrow act utilitarianism that treats common moral rules as mere shorthand in the service of moral cost–benefit calculation must be wrongly invoked in the affirmative answers to the moral 'dilemmas'.

[9] See also my review of Nathanson's book in *Social Theory and Practice* (2012). I am ignoring here the distinction between utilitarianism and consequentialism because it is also widely ignored by the neuroscientists. I think my criticisms of the understanding of utilitarianism and the implications of that would be readily adaptable to talk of consequentialism. Like the neuroscientists, and their philosophical supporters, I will use the term 'deontologist' to cover all forms of non-utilitarian/non-consequentialist ethical theories, though this is not strictly correct.

138–46).[10] It is hard to see her reasons for that as the result of an erroneous reliance on emotion, whatever her brain states are doing. Here I am not defending such deontological theories and theorists but merely pointing out the complexities that are consistently ignored (or just not understood) by neuroscientists. A further point is that utilitarians, unlike most non-utilitarians, have no room for a ground-floor distinction between the permissible and obligatory, so responses to what is permissible or non-permissible (which most of the experiments call for) inevitably leave it open whether the permissible response is a (sloppy?) shorthand for obligatory or carries a non-utilitarian message.[11]

One might also note that the loose vocabulary of 'moral dilemma' (used by all the psychological studies of moral decision-making) obscures the issues in another way—namely, that philosophers tend to use the term to denote those situations in which there is no right or wrong answer to the decision problem that the agent faces. In the examples considered by Bartels and Pizarro, different agents give different answers to the puzzling cases, but each thinks that their answer is unequivocally right without moral remainder. Or if they don't actually think that, the experimental scenario forces them to act as if they do. For dirty hands theorists the story they tell (whether coherent or not) is that the action is morally wrong, but nonetheless the right thing to do; for moral dilemma theorists either choice of actions is neither right nor wrong—morality drops out at this point. It may be thought that this qualm is too technical to insist upon, but I think the point is worth making because it not only highlights the philosophical misunderstandings of many neuro-scientists, but also because it shows the extreme artificiality of the demand for a binary response (yes/no, permitted/not permitted) in the use of the dilemma examples.

Joshua Greene, for instance, who has a philosophy background and should not be ignorant of such complexities, cheerfully treats various extreme philosophers' examples, broadly characterized as 'trolleyology', as being 'moral dilemmas [that] reveal the structure of moral cognition. They are moral illusions—revealing for the manner in which they mislead us' (Greene 2013, p. 252). The misleading is invariably done by a 'moral gizmo' in the brain that produces emotional reactions as 'alarm bells' that warn us off various acts, such as pushing an innocent bystander under a trolley. Although Greene's ruling picture of such gizmos is of moral illusions (comparable to such visual illusions as the Müller-Lyer lines) that must be subjected to the corrective power of utilitarian rationality, he occasionally admits that they can do some good and maintains only that they are not 'infinitely wise' (Greene 2013, p. 253). But then neither is cool reason (nor any other faculty or brain function), as such brutal paragons of unemotional, cool rationality as Victor Hugo's dedicated policeman, Javert (in *Les Miserables*), forcefully illustrate. Javert's stern commitment to his duty to enforce the

[10] She has elaborated her deontological reasoning in a critical response to Greene and others in her paper (Kamm 2009). She also criticizes the 'personal/impersonal' story.

[11] Cf. Kamm, pp. 339–40.

law against the noble escaped convict Valjean is eventually and reluctantly overcome by his emotion of gratitude at Valjean's rescuing him from death at the hands of revolutionaries, but such is his distress at his sense of betrayal of moral rationality that he kills himself.

Defects in the laboratory studies of our 'moral brains' extend beyond the considerations I have adduced so far to the likely irrelevance of such studies to the realities of moral thinking and emotion in the world where they actually matter. Not only are the binary responses inadequate reflections of the philosophical theories they purport to reflect but, as philosophers such as Appiah, Berker, and Glannon have pointed out, the scenarios considered in laboratory conditions are unlikely to elicit the thinking and feeling that real-life encounters with difficult moral decisions will involve (Appiah 2008; Berker 2009; Glannon 2011). Being asked to imagine how you would respond, and why, to a described emergency scenario is not the same as being in that scenario, and it is a big assumption to think a) that the reasoning and feeling that occurs in the former will be the same as that in the latter, and b) that the brain states involved in both will be the same. There are many reasons to doubt any such identities. One is that the real-life scenario will very probably differ in myriad detail from its imagined twin, thus giving rise to options deliberately excluded from the twin. Another is that my engagement with the real case confronts me with actual people and circumstances, and not descriptions or even visual presentations in a laboratory; there is little reason to think that my responses to, and guesses about, the latter will match my responses and thought about the former. It does, however, further highlight the lack of philosophical sophistication and intellectual depth involved in the way neuroscientists often resort to philosophical examples and theories.

Returning to Bartels and Pizarro, I think that in spite of some of their work involving certain of the confusions discussed above, a number of their points survive. No doubt they have drawn on philosophical positions too crudely and inaccurately, and been ignorant of much of the complexity in philosophical debates, but at least they have shown that a narrow version of utilitarianism (or consequentialism) that is at least a widespread tendency of thought, influential upon many in the wider community as well as intellectuals, should not be uncritically assumed by experimenters properly to characterize genuine moral thinking and deciding. Moreover, their experiments, though sharing some of the problematic artificiality complained of above, do strongly suggest that the casting of all emotions involved in moral thinking and deciding as error-inducing by the contrast with a certain model of rationality is ill-considered. The authors are also clear that they do not mean to imply that all people who have thought their way through to an endorsement of utilitarianism as a philosophical theory of morality are thereby exhibiting, or prone to, psychopathy. Furthermore, when their research is purged of misleading assumptions, it does suggest ways that the concentration on an emotionally detached model of reason may itself produce 'errors', since it can obscure or even omit altogether the possibly benign

role of emotions, such as empathy, in reasoning and thinking about moral questions, and thereby fail to acknowledge the dangers that adherence to that model poses for a fulfilling life.

2.4 The Nature of Emotions

The term 'emotion' is something of a portmanteau word, gathering in as it does a wide range of phenomena that are sometimes as different as they are similar. Rather than attempt a tight definition, I want to deal with some significant examples that exhibit features that make the divide I began with seem less plausible. I will, however, begin by mentioning some features that are common to many of the things called emotions. The first, noted long ago by William James, is that they have more or less strong connections with bodily reactions and feelings. Anger is a prime candidate for emotion, and angry people typically become heated and their faces flushed; they experience some level of desire to behave aggressively to the object of the anger, and this is difficult to inhibit. Fear, another prime candidate, tends to produce apprehensive reactions, such as sweating, rapid breathing, or a feeling of coldness. Of course, there is a great range of individual variation in these associated feelings and reactions, but there would be something very strange about anger or fear that produced none of these.

The second point, however, is that an emotion is more than the associated feelings and so forth. Subjectivists and emotivists like to describe their views as non-cognitivist with the implication that emotion is as raw as its associated feelings and has nothing to do with the intellect. That seems doubtful in several directions. For one thing there are many emotions that have cognitive structure related to what is appropriate for them to be directed at. Fear must typically be governed by the thought of an object that is considered dangerous to the perceiver or agent. There are no doubt nameless fears or a general fear of the world, but at its core fear is governed by a fairly specific thought that concerns danger to the agent or to those closely identified with the agent. Where the thought of danger is totally misplaced, it is right to call it irrational even though in terms of its associated feelings and bodily effects it is still a sort of fear. (Dickens catches this well with his description of David Copperfield, safe on land, watching a huge storm blow up at sea: 'my whole frame thrilled with objectless and unintelligible fear'.) Similarly, anger is typically geared towards slights or injuries related to the agent's concerns for herself or others. It makes no sense to describe someone as angry at being well treated or at the successes of a close friend unless the agent thinks the good treatment is a piece of condescension, or that the friend's successes have somehow been achieved at some resented cost to the agent. Again, gratitude cannot coherently be felt for something that can in no way be understood as a gift. It is no accident or mere social prejudice that nameless fear or objectless anger are pathological. It is facts like these about emotions that have led some thinkers to virtually assimilate emotions to beliefs or judgements. Robert Solomon, following Jean Paul Sartre to some extent, is

pre-eminent in this venture, and those like him are often called 'cognitivists' about emotions (Solomon 2003).

The extreme cognitivist account fails to give enough weight to the accompanying feelings and sensations of many emotions and also tends to assimilate emotions too closely to beliefs. The latter point is important because the intentionality aspect of emotions is more akin to that of perception so that (as with perception) it is possible for an emotion to represent a state of affairs that the agent believes does not obtain. So just as an agent can see an object as elliptical but correctly believe that it is round, so she can be afraid of a spider, thus emotionally representing it as dangerous when she may believe on good evidence that it is harmless. We could, however, use the term 'judgement' broadly enough to include such representational content even where the agent doesn't endorse what is represented as a belief. In that sense, emotions (or at least many of the central emotions) involve judgement.

However, quite apart from the agent's thought or judgement about the object or the way it appears in emotion to them, there remains the question of whether the emotion really fits its object. The person who is afraid of something that is not really fearful or dangerous (as someone who is afraid of a realistic toy spider) has been led astray by the emotion, just as someone who misperceives in dim light their own reflection in a mirror for another person has been led astray by their senses. Similarly, someone who is indignant about a wrong that isn't really a wrong has an emotion that fails to match its object, even if the emotion is triggered by a relevant thought—namely, that of a wrong. There is also a sense in which people can still be led astray by their fear or indignation even when they know the spider is a toy or the supposed affront is not a wrong.

Emotions are also related to the way the world is in a forward-looking fashion in that they prompt a typical range of object-related reactions such as flight from the danger feared or more positive ways of eliminating it. Admiration prompts emulation, indignation prompts attempts to remedy the wrong or injustice, and jealousy prompts the sort of actions that led to Othello's disastrous decline.

A further thing about emotions that invokes a cognitive or intellectual dimension is the way they can be discordant or harmonious with each other and with the agent's quest for a fulfilling life. One does not have to have a rigid conception of what such a life must be like, but there are limits to what it can be. Emotions such as envy, jealousy, spitefulness, and (with some qualifications) hatred carry with them disturbances of emotional balance that interfere with such emotional dispositions as kindliness, love, admiration, and gratitude.

Any attempt to live a good life must at least involve a modicum of attention to having good reasons for what we do and how we plan our lives (however minimally we do that planning). If we gear rationality to living rather than, or as well as, theorizing, we must pay particular attention to the rational dimensions of emotion, and not view emotions as mere raw, distorting turbulences. Indeed, there is good reason to think of trusting our emotions in the same sort of way that we trust our other resources for dealing with

the world we live in. What we usually consider to be our strictly cognitive resources for recognizing and exploring the world—for example, sensory perception, memory, introspection, inferential capacities and testimony—can all lead us astray, as can our emotions, but we need to have a general trust in those cognitive resources, not only to survive but to have any prospect of living well. Epistemologists have increasingly realized that attempts to somehow 'prove' the reliability of such cognitive resources without circularity are doomed to failure. This, of course, is the crucial move in sceptical arguments, but, assuming that scepticism has little to tell us about how to live (and here I side-step certain ways in which ancient sceptics did indeed think that they had provided a lesson for living), what the circularity problem helps to show is that there is a certain trust in our cognitive faculties that is central to coping with ourselves and the environment (whatever ontological status that environment ultimately has). I want to argue that the same point applies, *ceteris paribus*, to our emotional capacities.

Linda Zagzebski has made the case that trust in our cognitive resources for understanding and coping with the world is not only the upshot of the futility of trying to prove their reliability but is connected to a basic form of self-trust (Zagzebski 2012). Her point is that self-consciousness involves a natural inclination to seek truth, to pose questions and seek correct answers to them, even if this natural desire is often at work prereflectively. The workings of this desire at both the prereflective and the reflective levels requires also a natural desire to remove what she calls dissonance both between beliefs and between beliefs and desires, or feelings, or actions. The existence of such dissonance and the need to remove it is what leads often enough to what she calls 'conscientious self-reflection', which is a key notion in her concept of rationality. It is by conscientious self-reflection that the self seeks to resolve dissonance whether this involves beliefs, actions, desires, or emotions, or some combination of them. But underpinning the operation of such self-reflection is basic trust in the self and in the self's natural resources for seeking truth. Rationality consists, for her, in 'doing a better job of what we naturally do' (p. 45). Put differently, her argument is that trust in our basic cognitive faculties is not a default position waiting upon the failure of independent non-circular attempts to justify them but is basic to anything we can understand as a rational enterprise.

Although a principal concern of Zagzebski is with our cognitive capacities in the rather narrow sense common in traditional epistemology, she thinks that her argument applies to our desires and emotions as well. As she puts it, 'I have the same general grounds for trust in my emotion dispositions as I have for trust in my epistemic faculties' (p. 93). She acknowledges the cognitive dimensions of emotion already discussed above and her argument from self-trust is intended to apply to them as well. Of course, this does not mean that we blindly trust all our emotional reactions in every situation in which they occur, nor do we blindly trust every piece of mental arithmetic, every visual inspection, every memory, or every piece of testimony. What we eventually trust is the outcome of the process of conscientious reflection that applies to resolve dissonance

in both the epistemic and the emotional cases (and the desire cases, which I will not here discuss).

The basic question to which trust in emotions is addressed is whether they genuinely fit their objects. But there are other questions about them up for adjudication, such as whether some emotions are oriented to the wrong reactions to their objects. This concerns the issue of what patterns of action an emotion 'prompts', as I put it earlier. Emotions can be cognitively apt for certain objects but produce a distorted picture of the place of those objects in the world we inhabit. Consider envy. The appropriate object for envy is (very roughly) the successes or achievements of another; except for very unusual circumstances, we cannot envy failure in another, nor can we envy a rock (though we might envy a neighbour who has secured a particularly fine-looking stone for a centrepiece of their garden). But, of, course the success of another is also the proper object of the emotion of admiration. Comparing the achievements or high qualities of another person with our own is integral to the generation of both envy and admiration. But the reactions of those emotions go in opposite directions. Admiration leads naturally to emulation in some degree, or at least to satisfaction that a fellow human being (colleague, co-national, co-religionist, etc.) can achieve so well, and this affords a certain encouragement to oneself. Envy, on the other hand, leads to disparagement of the other's achievements, and often it stimulates thoughts that the other's success or honour is undeserved. When we say to someone, 'You're just envious of Smith', we mean to highlight the fact that the emotion of envy has distorted the agent's perception and understanding of the achievements of the envied person. Typically that is what envy does: it misrepresents aspects of the world to the agent, not, of course, of the facts of the other's achievements but of the way they were achieved, or of their true significance. If one's disparaging thoughts were actually true—that is, for example, if the person's successes were not due to their personal excellences or relevant hard work, but to toadying or bribery or something similar—then one's reactions should not be envy but rather a healthy emotion, such as indignation or something of that sort.

Of course, this is just a sketch of the territory of rationality and the emotions, and there are many more details to sketch, and more shades and moderating touches to be made. One point that immediately arises is the question of whether the trust in self that underpins trust in the emotions can reach as far as it does in the case of trust in the traditionally epistemic faculties. I am not sure of the answer but I will briefly explore some of what seems to me at issue. The question has at least two dimensions. The first concerns the diversity of emotions in two respects (1 and 2 below) and the second concerns the morally significant ways in which emotions may miss the mark (3 below):

1. It is claimed, for instance, that some emotions apparently central to morality are more local than we like to think. This has been argued with respect to what seems an emotion central to morality—namely, guilt. There are other cultures,

so it is said, that experience shame but have little or no place for the emotion of guilt, or treat shame as central to 'ethical behaviour' in a way that is more significant than the modern concentration on guilt. Bernard Williams is a notable philosopher who has discussed this sort of case by comparison of the ancient Greek outlook with that of post-classical Western cultures (Williams 1994). I am not entirely persuaded by his argument, or this particular example, but it has some plausibility and needs discussion. It must however be borne in mind that arguments from cultural variation have their pitfalls. The fact that some cultures have had no place for natural science as we understand it or for historical study as we understand it is no objection to the rationality of those endeavours even if we might learn something from what was in their place.

2. Individuals may vary more in their emotional capacities than in their epistemic. Individuals do, however, vary considerably in their intellectual and perceptual capacities, and any given individual does so throughout their life, as I can personally testify about my ageing memory and my increasingly poor hearing. But maybe the variations are more significant with emotions: the capacity to be amused, the emotion of sympathy (on which David Hume built his account of morality), the emotion of anger, the sense of the ridiculous, the emotion of fear—all these seem to vary a great deal amongst people who appear otherwise healthy enough.
3. It seems possible that emotions may miss the mark (i.e. fail to match their appropriate objects) more frequently than the epistemic faculties. At least some of our emotions may be less stable than the faculties of sight, hearing, inferring, and so on. The emotion of disgust, for example, is contentious in many debates about bioethical innovations, precisely because we have come to see that cultural conditioning around this emotion has been prominent in creating, or at least sustaining, evils such as racism and homophobia. Disgust may nonetheless be morally and rationally important rather than a mere visceral reaction since feeling disgust at pictures of torture, or vivid accounts of war crimes or sex slavery, can help change radically for the better the way people think about social and political policies. Nor should discussion of the complexities of disgust and other emotions neglect the way that emotion can be rightly influenced by imagination, context, and thought. In the very advanced art gallery in Hobart, Tasmania, called the Museum of Old and New Art (MONA), there is an installation called a Poo Machine that mimics the operations of the human digestive system from the ingestion of food to its excretion. My grandson Samuel, then a rather sophisticated youngster aged 9, was watching this phenomenon with interest and when the machine deposited a life-like excretion, a lady standing near him cried out: 'That's disgusting!' To this Samuel replied sternly: 'No it's not. We do that and we're not disgusting.' No doubt reactions might have been different had they both beheld a visitor to MONA divest the relevant clothing and mimic the machine. So context, circumstances, and mindset feed into the

generation of disgust in a way that complicates the picture of it as a straightforward spurt of emotion. Perhaps some emotions can go astray more easily than others (contrast disgust with sympathy), but we should not lose sight of the fact that the traditional epistemic faculties have their pathologies as well. Vision is a broadly reliable way of finding out about our environment but it is subject to many misjudgements, as the evidence of skilful magician tricks and common optical illusions shows.

I hope I have charted some of the pitfalls in the tendency to think of our emotional life as detached from, largely hostile to, or in need of rigorous control by an attenuated reason, and to suggest that too much of the neuropsychological investigation of the role of reason and emotion in ethics attends too little to the subtleties and complexities of both philosophical ethics and psychological reality. There are two lessons for the enhancement project in this. The first is that the findings of those psychological cum neurological investigations that often go under the label of neuroscience should be scrutinized carefully and used with caution. The second is that various proposals to enhance either our reasonings or our emotions need to attend to the often intimate connections between reason and emotion since enhancing what we imagine to be one separate and superior category or subcategory is likely to have serious implications for the other.

References

Appiah, K. A. (2008). *Experiments in Ethics*. Cambridge, MA: Harvard University Press.

Bartels, D. M. and Pizarro, D. A. (2011). 'The Mismeasure of Morals: Antisocial Personality Traits Predict Utilitarian Moral Preferences', *Cognition* 121: 154–61.

Berker, S. (2009). 'The Normative Insignificance of Neuroscience', *Philosophy and Public Affairs* 37(4): 293–329.

Coady, C. A. J. (2012). 'Review of Stephen Nathanson', *Terrorism and the Ethics of War*, *Social Theory and Practice* 38(3): 560–7.

Damasio, A. (2007). 'Neuroscience and Ethics: Intersections', *The American Journal of Bioethics* 7(1): 3–7.

Damasio, A. R. (1995). *Descartes' Error: Emotion, Reason and the Human Brain*. London: Picador.

Deonna, J. A. and Teroni, F. (2008). *The Emotions: a Philosophical Introduction*. Abingdon: Routledge.

Glannon, W. (2011). *Brain, Body and Mind: Neuroethics with a Human Face*. Oxford: Oxford University Press.

Greene, J. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York: Penguin.

Held, V. (2005). *The Ethics of Care: Personal, Political and Global*. Oxford: Oxford University Press.

Kamm, F. (2007). *Intricate Ethics: Rights, Responsibilities and Permissible Harm*. Oxford: Oxford University Press.

Kamm, F. (2009). 'Neuroscience and Moral Reasoning: a Note on Recent Research', *Philosophy and Public Affairs* 37(4): 330–45.

Kehane, G., Everett, J. A. C., Earp, B. D., Farias, M., and Savulescu, J. (2015). '"Utilitarian" Judgments in Sacrificial Moral Dilemmas Do Not Reflect Impartial Concern for the Greater Good', *Cognition* 134: 193–209.

Lloyd, G. (1984). *The Man of Reason: 'Male' and 'Female' in Western Philosophy*. London: Methuen.

Long, A. A. and Sedley, D. N. (eds) (1987). *The Hellenistic Philosophers*. Cambridge: Cambridge University Press, 65F, Diogenes Laertius 7.116.

Mill, J. S. (1957 [1873]). *Autobiography*. Indianapolis and New York: The Liberal Arts Press.

Nathanson, S. (2010). *Terrorism and the Ethics of War*. Cambridge: Cambridge University Press.

Nussbaum, M. (2001). *Upheavals of Thought: the Intelligence of Emotions*. Cambridge: Cambridge University Press.

Ross, W. D. (2002). *The Right and the Good*. Oxford: Oxford University Press.

Ruddick, Sara (1989). *Maternal Thinking: Towards a Politics of Peace*. Boston: Beacon Press.

Slote, M. (2007). *The Ethics of Care and Empathy*. New York: Routledge.

Solomon, R. (2003). *Not Passion's Slave: Emotions and Choice*. Oxford: Oxford University Press.

Tversky, A. and Kahneman, D. (1981). 'The Framing of Decisions and the Psychology of Choice', *Science* 211(4481): 453–8.

Williams, B. (1994). *Shame and Necessity*. Berkeley: University of California Press.

Zagzebski, L. T. (2012). *Epistemic Authority: a Theory of Trust, Authority, and Autonomy in Belief*. Oxford: Oxford University Press.

3

Repugnance as Performance Error

The Role of Disgust in Bioethical Intuitions

Joshua May

3.1 Arguments from Repugnance

Disgust inevitably crops up in bioethics since it is connected with our attitudes towards the purity of the body and mind (Nussbaum 2004; Haidt 2012). Such concerns are relevant to human enhancement, which can involve manipulating the building blocks of human life and chemically altering the brain. Leon Kass (1997, 2001), for example, has urged that altering human nature, especially by human reproductive cloning, is repellant, grotesque, Frankensteinian, revolting, and repugnant. He even compares this to disgust toward incest and cannibalism. He does offer other arguments that do not directly appeal to repugnance, but these seem secondary, proffered only so that this reaction is 'at least partially articulated' (Kass 2001, §3).

Others also endorse appealing to disgust in moral arguments (e.g. Miller 1997; Kekes 1998; Kahan 1999), but concerning other topics, such as sexual conduct and vicious cruelty. Given their usual conclusions, such positions are sometimes considered conservative. This is not true of all (e.g. Kahan 1999; Midgley 2000), but appealing to disgust might also be considered conservative given that empirical evidence suggests that political conservatives are more sensitive to disgust than liberals (Inbar, Pizarro, and Bloom 2009).

Arguments from repugnance rely on assumptions about the psychology of moral judgement. One is that such reactions toward bioethical issues are widespread. Kass (1997, p. 21), for example, refers to 'the widespread repugnances of humankind' (cf. also Kekes 1998, 100ff). After all, the argument would lack force if the reactions occurred only in the person making the argument or a small group of disgust-sensitive theorists. Data I report elsewhere challenge this initial assumption (May 2016), but here I want to focus on the second assumption, which is that emotions like disgust play a significant role in generating such bioethical judgements. This is an important claim because, even if disgust is a common reaction to certain uses of biotechnologies, it may merely be a byproduct of moral judgement. Opposition to cloning, for example, could

primarily be determined by factors other than repugnance, such as fear or thoughts about harm and fairness. If disgust is a mere consequence of an existing negative moral belief, however, then it cannot be cited as evidence for the truth of the moral belief. Arguments from disgust are meant to provide evidence of immorality, not evidence of a moral belief one already possesses (cf. Giubilini 2015). Thus disgust advocates must assume that the emotion drives the bioethical judgement, not vice versa, and that this influence isn't negligible.

I shall argue that recent work in cognitive science challenges this second assumption. Moral cognition is at best only slightly influenced by feelings of repugnance, and this minor effect should be treated as yielding a 'performance error'—akin to judgements influenced by fatigue, divided attention, agitation, and so forth. Disgust is thus not a factor involved in the normal production of moral judgement (it's 'exogenous'), failing to even play a role in our ordinary moral understanding, let alone moral wisdom. On this account, repugnance is certainly not 'above all [...] a moral and social sentiment' (Miller 1997, p. 2).

This yields a new kind of challenge to arguments from repugnance. Some commentators quickly dismiss appeals to emotion generally, arguing that such appeals are guilty until proven innocent by the tribunal of reason (e.g. Pence 1998; cf. Roache and Clarke 2009). Others have argued positively that disgust is morally irrelevant or unreliable in such contexts (e.g. Nussbaum 2004; Kelly 2011). My argument ultimately turns on the more basic premise that, for moral judgement, disgust is *psychologically* irrelevant. Such an account is insulated from replies from disgust advocates who argue that the emotion is morally relevant in some circumstances (e.g. Plakias 2013).

In the end, even if it is sometimes wrong to violate the purity of the body, this is not supported by appeals to repugnance. Our focus throughout will be on disgust and human enhancement, but the approach can obviously be applied to arguments from disgust on any topic, such as homosexuality. Moreover, we will briefly conclude that there are some broader implications of treating disgust as generating mere performance errors.

3.2 Disgust Experiments

Many say we can safely conclude that disgust substantially influences certain moral beliefs (e.g. Miller 1997; Prinz 2006; Kelly 2011; Haidt 2012; Chapman and Anderson 2013; Plakias 2013). Recent empirical research apparently suggests that moral judgement is largely driven by emotions, and many studies involve disgust in particular. Daniel Kelly, for example, proclaims that disgust 'can have dramatic effects' on moral judgement (2011, p. 130). Similarly, Chapman and Anderson write that numerous studies 'converge to support the notion that disgust does play an important role in morality' (2013, p. 322). I will argue that such treatments of the evidence are overblown. It is imperative to begin with an examination of some of the key studies.

Mere correlations between disgust and moral judgement are perhaps commonplace and uncontroversial, at least for some moral judgements. So let's grant that there is

sufficient evidence for the correlations, whether based on experiments or common experience (cf. Prinz 2006). Still, it could be that disgust is merely a result of the moral judgement or otherwise does not play a causal role in its production. Often it seems we are disgusted by something because we judge it to be immoral (not vice versa), and there is some neuroscientific evidence for this ordering (Yang et al. 2013). Emotions are commonly triggered by judgements of wrongdoing, presumably because morality matters a great deal to us. So let us turn to evaluating studies that attempt to isolate disgust as a causal factor in moral judgement.

Consider first the popular 'moral dumbfounding' studies (e.g. Haidt, Koller, and Dias 1993). After subjects recorded their moral judgements about various 'harmless taboo violations', such as eating an already dead dog, the experimenter probed for justifications using an interview format. But participants who condemned the acts were at a loss. They tended to look for harms in the scenarios only to be reminded that there apparently weren't any in the cases under consideration (but see Jacobson 2012). Participants were often in a state of moral dumbfounding—confident in their beliefs but uncertain why. Haidt and his colleagues argue that this phenomenon further reveals that emotion is a key determinant of the initial moral judgement, and 'moral reasoning is usually post hoc rationalization' (Haidt and Bjorklund 2008, p. 216). What's most important for us is that many of the hypothetical scenarios involved violations of the purity of the body, and the researchers found a correlation between how negative participants' moral judgements were and reports of being 'bothered' by the scenarios. It would thus seem that disgust is the relevant emotion providing the causal influence, at least in these types of scenario.

Such studies might actually seem damaging to arguments from repugnance. They may suggest that disgust influences moral judgement in an irrational way—hardly providing wisdom. But that's too quick. Haidt (2012) himself maintains that disgust plays a powerful role in moral judgement, and one that we should take seriously. The culprit is not emotion but rather conscious reasoning, which engages in post hoc rationalization. On Haidt's theory, roughly, disgust is tied to one of his hypothesized innate moral foundations—namely, sanctity/degradation—that he believes evolved in humans partly to shape moral cognition regarding topics concerning the purity of body and soul. At one point, Haidt goes so far as to say: 'If we had no sense of disgust, I believe we would also have no sense of the sacred' (p. 349). This is fodder for Kass and other disgust advocates. At one point, Kass himself seems to welcome the idea that 'repugnance is the emotional expression of deep wisdom, beyond reason's power completely to articulate it' (2001, §2). Similarly, Kekes writes: 'If challenged to justify their reaction, [those feeling moral disgust] may not be able to do so. But that does not mean that their reaction is not justifiable' (1998, p. 106).

Support for this picture may come from a general idea in cognitive science that a judgement is not suspect just because one cannot articulate one's reasons for it. For example, Noam Chomsky's famous sentence 'Colorless green ideas sleep furiously' strikes ordinary speakers as grammatical (or acceptable), but they cannot explain why;

we have to be taught why in grammar school or linguistics class. One can point to non-linguistic examples as well, such as our general inability to explain why we love someone (Saltzstein and Kasachkoff 2004). This phenomenon is unsurprisingly well documented for various moral judgements, even those that do not involve tricky 'harmless taboo violations' (e.g. Cushman, Young, and Hauser 2006). Proponents of the so-called 'linguistic analogy' argue that, much like linguistic judgement, moral judgement is generated by factors that are not easily accessible via introspection (e.g. Dwyer 2009; Mikhail 2011). So a range of theoretical frameworks can explain moral dumbfounding without treating disgust as unreliable.

However, the dumbfounding studies do not provide much evidence about whether repugnance does influence moral judgement. The main issue is that disgust wasn't manipulated experimentally or directly measured; participants were only asked how much the acts depicted 'bothered' them. Disgust *might* be the relevant emotion in many of the scenarios, but it's best to examine the various experiments that have induced disgust in participants and apparently observed a change in moral judgement due to an incidental experience of this emotion alone.[1]

There are indeed a number of such experiments purportedly isolating disgust as a causal factor in moral judgement. For example, Wheatley and Haidt (2005) manipulated disgust by hypnotizing subjects to feel disgust on hearing a morally neutral trigger word. Similarly, Simone Schnall and her colleagues induced disgust in their participants in a variety of ways, such as having them smell a foul odour, watch a film clip involving human faeces, and undertake the experiment in a gross environment (Schnall et al. 2008). In these and a number of other experiments, researchers have allegedly found evidence that incidental disgust substantially influences moral judgements—making them, in particular, more negative compared with those of control groups. A related set of studies involved manipulating feelings of cleanliness, but its effect on moral judgement is more complicated. Some results suggest that cleanliness reduces the severity of moral judgements, while others suggest an increase in harshness (for a review, see Tobia 2015).[2]

Various commentators, often including the experimenters themselves, have claimed the data suggest that those who experienced disgust were more likely to judge hypothetical actions as wrong or immoral, compared with a control group who were not induced to feel disgust. For example, regarding the experiments conducted by Schnall et al. (2008), Joshua Greene writes: 'the disgust manipulation made people more likely

[1] There are further problems with drawing firm conclusions about moral judgement from the dumbfounding studies, but we'll set them aside here. See, for example, Saltzstein and Kasachkoff 2004, Jacobson 2012, and Kennett 2012.

[2] I discuss other similar experiments, and their limitations, elsewhere (May 2014a). Haidt and colleagues do present additional evidence that emotions generally influence moral judgement (see especially Haidt and Bjorklund 2008). Much of the evidence, however, is speculative and does not directly concern disgust or moral judgement, such as evidence for dual-process theory (see Saltzstein and Kasachkoff 2004). The most important studies to evaluate involve experiments that manipulate disgust and measure moral judgement specifically.

to condemn these actions' (2008, p. 58). Based on such results, some philosophers have gone so far as to make claims such as the following: 'we can form the belief that something is morally wrong by simply having a negative emotion directed towards it' (Prinz 2006, p. 31).

3.3 Disgust's Influence Tempered

In previous work I argued that such experiments do not provide strong evidence for the claim that disgust substantially influences a surprising class of moral judgements (May 2014a). Here I want to briefly rehearse some key components of that argument, extend them to bioethics, and draw a bolder conclusion. This will set us up for addressing the arguments from repugnance with which we began.

The most important point about the disgust experiments is that they consistently show only that disgust *slightly* influences moral judgement. In these studies, moral judgements are importantly recorded on a scale, anchored at different end points with moral categories such as 'Not at all morally wrong' vs. 'Extremely morally wrong'. Participants are randomly assigned into groups—at least two, consisting of one disgusted group and one control group—and they all read a number of hypothetical scenarios, typically involving moral transgressions. The experimenters subsequently calculate the average response on the scale in each group and then use statistical analyses to determine whether the difference between the two groups is likely due to chance. That is, they engage in the common method of null hypothesis significance testing, assessing statistically whether we can confidently reject the null hypothesis that there is no real difference between such groups (any observed difference is just due to chance). Across the relevant experiments, the researchers tend to find that the mean response from the disgusted group is greater than the mean of the control group, and, importantly, that this difference is statistically significant. We can then reasonably reject the null hypothesis, concluding with some confidence that the manipulated variable (disgust) had a causal impact on moral judgement.

We must further ask, however, whether the difference is *substantial*, since statistical significance alone does not entail that the effect is significant in a more ordinary sense. The mean differences between the groups' responses may, after all, be ever so slight; and in fact they are, across the board. In Wheatley and Haidt's (2005) hypnotism study, for example, one of the differences in mean morality ratings was 2.7 and 14.0 on a 100-point scale. This, if anything, appears to be a rather insignificant difference. It is statistically significant to be sure, but it is not a substantial shift, at least as far as our topic of moral judgement is concerned. This does not appear to provide any evidence for the claim that participants' moral judgements changed in terms of their *polarity* or *valence* (e.g. from right to wrong). The means are both on the same side of the scale, suggesting that participants in both groups tended to register the same moral judgement (namely, *not wrong*) regarding the case (cf. Mallon and Nichols 2010, pp. 317–18). So we do not have evidence that moral judgements, even a specific class, can be driven

merely by feelings of repugnance (contra Prinz 2006 and others). Rather, the evidence suggests that disgust can slightly amplify moral judgements or make them harsher.[3] This is a limitation that continually arises for the many disgust experiments that have been conducted by various labs around the globe. Moreover, a recent meta-analysis of published and unpublished attempts to produce the effect suggests its magnitude is small at best (Landy and Goodwin 2015).[4]

The experimental results provide *some* evidence for a rather modest claim: that disgust *slightly* influences the *severity*, not the valence, of moral judgements in a certain domain. Such minor influence is akin to factors, such as fatigue or inattention, which slightly impact the severity of judgements generally. In fact, there is evidence that disgust's influence works much like fatigue. Experiments by Simon Laham and his colleagues (Laham, Alter, and Goodwin 2009) suggest that moral judgements can be made less severe by, in effect, reducing incidental fatigue, by varying the legibility of the font their moral vignettes were presented in. For one group of subjects the first few vignettes were difficult to read while the latter few were refreshingly easy, and the reverse was true for the other group. As with disgust, reduction in the fatigue that presumably accompanies illegible writing slightly affected the severity of people's moral judgements on average, but not their valence. The mean difference in morality ratings between the groups was statistically significant of course, but the means themselves differed only slightly and on the same side of the scale.

Such findings are not particular to morality either. For example, there is some evidence that claims about geography, of the form 'Town A is in county B' (e.g. 'Lima is in Peru'), are slightly more likely to be judged true when more legible (Reber and Schwarz 1999). More precisely, the group of participants who read the highly visible statements judged slightly more statements as true (8.36 statements) than the group that read the same statements presented in a less visible colour (8.09 statements). The difference between these means is statistically significant, but one can readily see that the difference between the groups is slight. Such subtle effects have been documented on many topics; there is nothing peculiar about geography or morality.

This might seem like a wholesale attempt to water down the results of all psychological experiments, but it is important to note that the limitations we've noted are not criticisms of the studies themselves and they do not apply to all of them. Certain projects are geared towards finding any shifts whatsoever on the relevant instrument of measurement. To take a simple example, one might be concerned to measure helping behaviour in different conditions (cf. Batson 2011). If we measure this behaviour by documenting the percentage of people who help when, say, feeling or not feeling especially high levels of empathy, then any statistically significant shift can yield results with clear importance. One key difference between studying helping behaviour and

[3] In addition to May 2014a, this issue is raised briefly by Huebner, Dwyer, and Hauser 2009, Pizarro, Inbar, and Helion 2011, and Royzman 2014.

[4] In order to focus on what is most important for the argument in this chapter, I am setting aside further limitations of the various disgust experiments. See May 2014a and Huebner 2015.

moral judgement is that the units of measurement for helping behaviour are each rather substantial: each percentage point represents a portion of people helping. With standard scales for moral judgement, however, each unit of measurement represents at best something like a tendency to shift one's belief about *how* moral an act is, or perhaps a shift in average confidence in the categorical judgement. So, as we've seen, even if there is a statistically significant difference in the average responses between groups, this may still represent the very same judgement (e.g. that the action is morally acceptable). So both the goal of one's project and the phenomenon measured can affect whether statistical significance alone yields something of immediate importance for the research programme.

A primary goal in research on moral judgement is to isolate which factors determine the valence of moral beliefs. In the arguments from disgust at least, the idea is supposed to be that disgust is in some sense a major determinant of the judgement. In appealing to repugnance, bioethicists such as Kass are in effect claiming that they believe some action or policy is wrong because they have a reaction of repulsion upon contemplating it. Such appeals are impotent if disgust only slightly affects the harshness or severity of the judgement, not its valence or polarity. Compare the following: one's mood may slightly affect the severity of one's judgement that Jones is a jerk; but one could hardly appeal to one's mood as indicating that Jones is a jerk if one would believe this regardless of one's mood. Alternatively, consider not a judgement but a mere cause-and-effect relationship: adding insult to injury. Suppose Leslie falls into a bout of depression because she is recalled from her beloved job as a City Council member. The knowledge that she lost her job is the main cause of her being depressed (rather than happy), even if literally insulting her would make her depression *worse*.

3.4 Performance Errors and Exogenous Factors

So far we have considered in rather ordinary terms the limits of disgust's influence on moral judgement. But we can appreciate these limits more fully with the so-called 'competence/performance' distinction used in many areas of cognitive science.[5]

The distinction was made most prominent by Chomsky's approach to language, which distinguished between one's understanding (*competence*) and how one puts it to use (*performance*). *Performance errors* arise not whenever one makes a mistake about the subject matter but rather when one's performance fails to reflect one's understanding. To take a simple example of a linguistic performance error, consider a case in which someone says, 'You did good', after seeing someone else successfully complete a difficult task. Some adult speakers of English may not know that one should say 'well' instead of 'good' in such cases, but we can sometimes make this 'error' even when we

[5] My thinking on this topic has been greatly influenced by the insightful work of John Mikhail (2011). He very briefly connects emotions and performance errors in his commentary on Greene's work (Mikhail 2008, n. 5).

do have the relevant knowledge. For example, this may occur when one is especially tired, in a hurry, distracted, or simply in the mood to be colloquial. So this error in performance—a slip of the tongue, we might say in some cases—need not reflect one's knowledge, or minimally beliefs, about the appropriate use of 'well'.

Upon a bit of reflection, one might readily notice performance errors, but others are not easily shaken. Consider the phenomenon of centre-embedding, in which part of a phrase is embedded in the middle of another. Single-embedding often seems grammatical (e.g. 'The salmon that the dog chased fell'); yet double-embedding is often unacceptable (e.g. 'The salmon that the man that the dog chased smoked fell'). Despite the second example sounding odd to ordinary English-speakers, many linguists chalk this odd ring up to performance error, perhaps due to a strain on memory (Lewis 1996; Alexander, Mallon, and Weinberg 2010). The judgement that the second example is unacceptable, then, needn't be indicative of our linguistic understanding (or 'competence').

Not all errors in judgement are performance errors. Some cognitive biases may be good examples of errors in judgement that arguably do reflect our intuitive understanding of the subject matter. For example, our tendency towards error when thinking about probabilities—such as the Gambler's Fallacy or the Monty Hall Problem—might best be treated as reflecting our basic, albeit erroneous, understanding of likelihood and chance. In such cases, the errors may well reflect our ordinary competence (in the technical sense), although 'competence' in the ordinary sense might not be the most appropriate term here since, in one sense, we are displaying *incompetence*. Similarly, talk of 'knowledge' in the ordinary sense is inapt insofar as knowledge entails accurate belief. But these are merely differences in terminology.

Indeed, the terminology here can court confusion, so let me be clear, especially since theorists sometimes use the competence/performance distinction in various ways. Some conceive of all actual psychological phenomena as a matter of performance, while competence is an abstract idealization that is not meant to describe actual psychological mechanisms (cf. Mikhail 2011). Moreover, some employ the notion of error in a more normative way to designate which judgements are incorrect or biased (cf. Nichols and Knobe 2007). A common role for the distinction in cognitive science, however, is merely descriptive—marking out the normal mechanisms in a domain of cognition (cf. Alexander, Mallon, and Weinberg 2010). Here the goal is to distinguish in a causal mechanism normal, internal factors (*endogenous*) from abnormal, external factors (*exogenous*). In moral judgement, examples of endogenous factors might be information about intent, act types, and harmful consequences (May 2014b), while an example of an exogenous factor might be fatigue. It may be better, then, to simply employ the distinction between endogenous and exogenous factors or variables.

One way to focus on this appropriation of the distinction is to consider the notion of a causal process via intervention (cf. Woodward 2003). A typical intervention into a causal system is a good example of an exogenous causal process. For example, suppose we normally acquire the belief that others are sad by processing information about

their facial expressions, utterances, and body language—the endogenous factors in this causal mechanism. One might, of course, form the belief that Sam is sad due to, say, being struck on the head by a falling brick. But this is typically an exogenous causal process that may reveal little about the endogenous causal mechanism that normally leads to reading the minds of others. Intervening on the brain may reveal much about how endogenous mechanisms work, by showing how they break down, but such interventions themselves are typically exogenous factors. Simply because such intervening factors can alter a subject's judgement, one would not conclude that such factors are part of the normal mechanism for producing such judgements. More evidence is required to make that inference.

Of course, some factors that we would intuitively categorize as interventions are endogenous factors or reveal such factors. But further evidence can help identify which interventions are endogenous or indicative of 'competence.' For example, in general, a causal factor is especially likely to be exogenous (possibly yielding 'performance errors') if it consistently fails to substantially affect the target variable. This may provide a sufficient, even if not necessary, condition for a factor being exogenous.

3.5 Disgust's Performance

What does the preceding discussion imply about disgust? If repugnance is best treated as an exogenous variable that yields 'performance errors' then its influence on moral judgement is even more limited than some other theorists have emphasized (e.g. Huebner, Dwyer, and Hauser 2009; Pizarro, Inbar, and Helion 2011; May 2014a). Not only does disgust only appear to slightly amplify moral judgements, rather than shift their valence, but it's not even part of the normal mechanism for moral judgement. If this is right then the influence disgust has on our moral judgements is like fatigue: its effects do not reveal our understanding of morality, which means it cannot constitute wisdom.

Of course, in the case of disgust we do not even have evidence of a shift in the categorical judgement of the rightness or wrongness of an action, which contrasts with the usual examples of performance errors. Instead we have evidence that the harshness of one's judgement can be slightly elevated. This is due to the scales used to measure moral judgement in these experiments. A slight shift along such scales is a change in response that does not necessarily reflect the ordinary mechanism. The valence of *some* people's moral judgements might be determined solely, or in large part, by feelings of repugnance. Kass is an example if we take his claims about his own psychology at face value. Perhaps some such people are especially sensitive to disgust or confuse their negative feelings of repugnance with, say, fear or anxiety. But, again, to have any dialectical force, arguments from disgust must appeal to the 'widespread repugnances of humankind'.

Now I do not have an a priori method for determining which factors are exogenous, but nor do I think one should. As some have pointed out, we should expect the competence/performance distinction to vary across domains (e.g. judgements about morality, language, etc.), and, partly because of that, we should expect the distinction

to be informed by the theory constructed in light of research on the particular domain. In this way, the distinction seems to be theory-dependent to some extent (Mikhail 2011, especially ch. 8). For example, for a non-linguistic case of performance error, some point to an experiment in which participants were asked, 'How many animals of each kind did Moses take on the ark?' The most prominent answer is 'two', despite the common knowledge that it was Noah, not Moses, who is supposed to have operated the biblical ark (Erickson and Mattson 1981). But notice that, prior to a well-developed theory, we already have some prima facie reason to count this as a performance error, given our knowledge about the topic and of people's occasional inattentiveness ('shallow processing'). At any rate, we can treat such responses as failing to reflect people's ordinarily understanding of this story.

The disgust experiments are similar in certain respects. The data gathered thus far suggest that disgust's influence is limited: when researchers do detect an effect, it amounts to a slight shift on a fine-grained scale. These do not, on the face of it, appear to be the kind of results one would expect from an endogenous variable. Consider factors that plausibly play a prominent role in ordinary moral cognition, such as beliefs about harm, fairness, rights, and so forth. These beliefs, and perhaps their attendant emotions, seem to have a substantial impact on moral judgement.

Consider, for example, an experiment conducted by Fiery Cushman (2008) which examined the effect of negative beliefs, desires, and consequences on moral judgement. Participants were randomly assigned to read variants of vignettes in which a negative outcome either did or did not occur and the agent either intended it to occur or not. In one scenario, Jenny is welding two pieces of metal together with a partner, intends for it to burn her partner's hand (she wants to and believes the welding will), and she succeeds. Subjects responded to questions about the wrongness of the relevant acts and the blameworthiness of the agents. Naturally, each factor had a statistically significant impact on judgements of each type. What's more interesting is that for judgements of wrongness (and permissibility) the agent's *belief* about the negative consequences of Jenny's action accounted for most of the variance in responses. In other words, Jenny's act was judged wrong largely because she knew what she was doing. Moreover, the mean response to this question shifted dramatically depending on whether or not the agent both believed and desired that the negative outcome would occur (roughly: if she intended it). When the agent intended the negative outcome in this way, participants on average appeared to condemn the act as wrong (even if it was unsuccessful and no harm occurred); yet the opposite was true when intentionality was absent. On a scale of moral wrongness, these two key factors (belief and desire) appeared to drive mean morality ratings from floor to ceiling, apparently altering the valence of judgements, on average. Given these data, and what we know about intention and moral judgement, we can readily count these factors as shedding light on the endogenous mechanisms of moral judgement, at least concerning harm. In fact there are plenty of such data confirming the common-sense idea that *mens rea* is an important part of moral cognition (for a review, see Young and Tsoi 2013; May 2014b).

Of course, the competence/performance distinction is not without its problems or detractors. For example, one might think a proper account of the distinction depends on having a correct theory of concepts in hand, which surely most of us lack (cf. Machery 2008). Others might insist that one possess a well-developed theory of the idealized cognitive system so that 'performance errors can be explained away in terms of the system falling short of that idealization in some way' (Alexander, Mallon, and Weinberg 2010, p. 305). However, we can make progress towards building a theory of moral judgement with more relaxed standards. Using a more bottom-up strategy, disgust can provisionally be treated as outside the normal mechanism of moral judgement by considering the data we have so far on its limited effects on moral judgement and comparing them with other kinds of influence.

Finally, notice that I have not hitched my project to the controversial 'linguistic analogy' and its Chomskyian framework. Proponents of the linguistic analogy (e.g. Dwyer 2009; Mikhail 2011) tend to believe that some amount of moral competence, even moral knowledge, is innate and universal, and they tend to conceive of moral cognition as arising from a module in the brain that is greatly insulated from other parts of the mind. One needn't make such claims in order to rely on something like the competence/performance distinction in a particular area of cognition (cf. Nichols and Knobe 2007 regarding judgements of responsibility). In studying the mind using responses from subjects, we are faced with the task of separating the wheat from the chaff when attempting to isolate the mechanisms that produce such responses from other factors and from the responses themselves.

3.6 Conclusion

Our focus has been on those theorists who base their bioethical conclusions in whole or in part on appeals to disgust. The argument from repugnance, however, relies on the increasingly popular empirical claim that disgust plays an important role in some areas of moral and legal thought. But this key empirical assumption is implausible. We do have some experimental evidence that disgust tends to *slightly* influence the *severity*, but not the valence, of some moral judgements concerning topics such as sanctity, purity, and degradation. The pro-disgust arguments, however, rely on a stronger claim: that repugnance substantially influences the relevant moral beliefs. The tempered picture of disgust's influence is importantly compatible with the idea, sometimes expressed by Kass and others, that we can sometimes be justified in our intuitively formed moral beliefs even if 'reason can't fully articulate' why. However, when such moral intuitions are influenced by disgust, we have reason to believe they are not reflecting the ordinary mechanisms producing moral judgement. Repugnance cannot serve as deep wisdom if it fails to reflect our moral competence or our intuitive understanding of morality.

Our conclusion might seem to conflict with common experience, given that disgust so commonly occurs with certain moral judgements. A quick search of the Internet provides a recent example in which author Philip Pullman called 'disgusting' a policy

that bans sending prisoners books (quoted in Flood 2014). But this use of 'disgust' is metaphorical, or just another way of saying that certain acts are not just wrong but especially heinous (cf. Gutierrez, Giner-Sorolla, and Vasiljevic 2012). (So it would be useless for disgust advocates such as Kass to switch to this or similar meanings.) Literal disgust is often just a consequence of moral judgement, not a cause. And, in those rare cases when it may seem to play a causal role, it is likely insubstantial.

This challenge to disgust advocates is rather distinct from others that focus on the moral irrelevance of disgust, such as those levelled by Nussbaum and Kelly. If I am right then the problem is even more basic. Still, my conclusion is in principle compatible with those of other 'disgust sceptics'. In fact, I tend to believe that they are best combined, posing a powerful challenge to those who appeal to disgust in making moral and legal arguments.

Let me close by noting a few broader implications of treating disgust as an exogenous factor in moral judgement. First, this challenge clearly applies to other moral arguments that appeal to disgust. Most notable are those against homosexuality and same-sex marriage, as discussed and critiqued at length by Nussbaum (2004), for example. I believe my account and hers make such appeals to disgust in the law and morality look rather dubious. Like fatigue, disgust is both psychologically and morally irrelevant.

Second, there is a problem for those who give disgust an important role in morality or the law. For example, Kahan, following Miller (1997), believes that 'disgust is an indispensible member of our moral vocabulary' (Kahan 1999, p. 64). Similarly, as already noted, Haidt (2012) proposes that one innate foundation of moral judgement involves intuitions about sanctity and degradation, and he importantly believes that disgust is the key emotion behind this. Such theories seem committed to the idea that disgust plays a role in moral competence, not merely generating performance errors. However, if I am right, then this is a mistake. It may even be a mistake to treat sanctity/degradation as a foundation at all, which casts doubt on a category hypothesized as especially important for conservative moral and political thought. At any rate, it at least seems that one cannot wed disgust to such a foundation; something else must be involved if it is to be part of moral competence.

A final implication: for all I have said here, this account is restricted to the emotion of disgust, but the treatment *might* be extended to other emotions. However, there have not been as many studies of the effect of incidental experiences of other emotions on moral judgement, so drawing any similar conclusions would be difficult at present.[6]

[6] Versions of this chapter were presented at the Southern Society for Philosophy and Psychology, Monash University, the University of Mississippi, the Georgetown University Philosophy Conference, and the Human Enhancement Workshop at the University of Melbourne. For especially helpful feedback on these occasions and others, I thank Josh Alexander, Monima Chadha, Steve Clarke, Alberto Giubilini, Dan Kelly, Christina Majoinen, John Mikhail, Gregory Pence, Walter Sinnott-Armstrong, and Aaron Zimmerman.

References

Alexander, J., Mallon, R., and Weinberg, J. M. (2010). 'Accentuate the Negative', *Review of Philosophy and Psychology* 1(2): 297–314.

Batson, C. Daniel (2011). *Altruism in Humans*. New York: Oxford University Press.

Chapman, H. A. and Anderson, A. K. (2013). 'Things Rank and Gross in Nature: a Review and Synthesis of Moral Disgust', *Psychological Bulletin* 139(2): 300–27.

Cushman, F., Young, L., and Hauser, M. (2006). 'The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm', *Psychological Science* 17(12): 1082–9.

Cushman, Fiery (2008). 'Crime and Punishment: Distinguishing the Roles of Causal and Intentional Analyses in Moral Judgment', *Cognition* 108: 353–80.

Dwyer, Susan (2009). 'Moral Dumbfounding and the Linguistic Analogy', *Mind and Language* 24(3): 274–96.

Erickson, Thomas and Mattson, Mark (1981). 'From Words to Meanings: a Semantic Illusion', *Journal of Verbal Learning and Verbal Behavior* 20: 540–52.

Flood, Alison (2014). 'Mark Haddon Launches Online Petition against Prisoners' Book Ban', *The Guardian*, 24 March, <http://www.theguardian.com/books/2014/mar/24/mark-haddon-online-petition-prison-ban-books>.

Giubilini, A. (2015). 'What in the World is Moral Disgust?', *Australasian Journal of Philosophy*, 4 August, doi: 10.1080/00048402.2015.1070887.

Greene, Joshua (2008). 'The Secret Joke of Kant's Soul', in Walter Sinnott-Armstrong (ed.), *Moral Psychology, Vol. 3: the Neuroscience of Morality.* Cambridge, MA: MIT Press, pp. 35–80.

Gutierrez, R., Giner-Sorolla, R., and Vasiljevic, M. (2012). 'Just an Anger Synonym? Moral Context Influences Predictors of Disgust Word Use', *Cognition & Emotion* 26(1): 53–64.

Haidt, J. and Bjorklund, F. (2008). 'Social Intuitionists Answer Six Questions about Morality', in Walter Sinnott-Armstrong (ed.), *Moral Psychology Vol. 2*. Cambridge, MA: MIT Press, pp. 181–218.

Haidt, J., Koller, S. H., and Dias, M. G. (1993). 'Affect, Culture, and Morality, or is it Wrong to Eat Your Dog?', *Journal of Personality and Social Psychology* 65(4): 613–28.

Haidt, Jonathan (2012). *The Righteous Mind*. New York: Pantheon.

Huebner, B., Dwyer, S., and Hauser, M. (2009). 'The Role of Emotion in Moral Psychology', *Trends in Cognitive Sciences* 13(1): 1–6.

Huebner, Bryce (2015). 'Do Emotions Play a Constitutive Role in Moral Cognition?', *Topoi* 34(2): 427–40.

Inbar, Yoel, Pizarro, D., and Bloom, P. (2009). 'Conservatives are More Easily Disgusted than Liberals', *Cognition and Emotion* 23(4): 714–25.

Jacobson, Daniel (2012). 'Moral Dumbfounding and Moral Stupefaction', in Mark Timmons (ed.), *Oxford Studies in Normative Ethics: Vol. 2*. Oxford: Oxford University Press.

Kahan, Dan M. (1999). 'The Progressive Appropriation of Disgust', in Susan A. Bandes (ed.), *The Passions of the Law*. New York: New York University Press.

Kass, Leon R. (1997). 'The Wisdom of Repugnance', *The New Republic* 216(22): 17–26.

Kass, Leon R. (2001). 'Preventing a Brave New World: Why We Should Ban Human Cloning Now', *The New Republic* 224(21): 30–9.

Kekes, John (1998). *A Case for Conservatism*. Ithaca, NY: Cornell University Press.

Kelly, Daniel (2011). *Yuck!: The Nature and Moral Significance of Disgust*. Cambridge, MA: MIT Press.

Kennett, Jeanette (2012). 'Living with One's Choices: Moral Judgment in Vivo', in Robyn Langdon and Catriona Mackenzie (eds), *Emotions, Imagination and Moral Reasoning*. New York: Psychology Press.

Laham, S., Alter, A., and Goodwin, G. P. (2009). 'Easy on the Mind, Easy on the Wrongdoer: Discrepantly Easy to Process Violations are Deemed Less Morally Wrong', *Cognition* 112(3): 462–6.

Landy, J. F. and Goodwin, G. P. (2015). 'Does Incidental Disgust Amplify Moral Judgment? A Meta-analytic Review of Experimental Evidence', *Perspectives on Psychological Science* 10(4): 518–36.

Lewis, R. L. (1996). 'Interference in Short-Term Memory: the Magical Number Two (or Three) in Sentence Processing', *Journal of Psycholinguistic Research* 25(1): 93–115.

Machery, Edouard (2008). 'The Folk Concept of Intentional Action: Philosophical and Experimental Issues', *Mind & Language* 23(2): 165–89.

Mallon, Ron and Nichols, Shaun (2010). 'Rules', in J. M. Doris and The Moral Psychology Research Group (eds), *The Moral Psychology Handbook*. Oxford: Oxford University Press.

May, Joshua (2014a). 'Does Disgust Influence Moral Judgment?', *Australasian Journal of Philosophy* 92(1): 125–41.

May, Joshua (2014b). 'Moral Judgment and Deontology: Empirical Developments', *Philosophy Compass* 9(11): 745–55.

May, Joshua (2016). 'Emotional Reactions to Human Reproductive Cloning', *Journal of Medical Ethics* 42: 26–30.

Midgley, Mary (2000). 'Biotechnology and Monstrosity: Why We Should Pay Attention to the "Yuk Factor" ', *Hastings Center Report* 30(5): 7–15.

Mikhail, John (2008). 'Moral Cognition and Computational Theory: Comment on Greene', in Walter Sinnott-Armstrong (ed.), *Moral Psychology, Vol. 3*. Cambridge, MA: MIT Press.

Mikhail, John (2011). *Elements of Moral Cognition*. Cambridge: Cambridge University Press.

Miller, William I. (1997). *The Anatomy of Disgust*. Cambridge, MA: Harvard University Press.

Nichols, Shaun and Knobe, Joshua (2007). 'Moral Responsibility and Determinism: the Cognitive Science of Folk Intuitions', *Noûs* 41(4): 663–85.

Nussbaum, Martha C. (2004). *Hiding from Humanity: Disgust, Shame, and the Law*. Princeton, NJ: Princeton University Press.

Pence, Gregory E. (1998). *Who's Afraid of Human Cloning?* Lanham, MD: Rowman & Littlefield.

Pizarro, D., Inbar, Y., and Helion, C. (2011). 'On Disgust and Moral Judgment', *Emotion Review* 3(3): 267–8.

Plakias, Alexandra (2013). 'The Good and the Gross', *Ethical Theory and Moral Practice* 16(2): 261–78.

Prinz, Jesse (2006). 'The Emotional Basis of Moral Judgments', *Philosophical Explorations* 9(1): 29–43.

Reber, R. and N. Schwarz (1999). 'Effects of Perceptual Fluency on Judgments of Truth', *Consciousness and Cognition* 8(3): 338–42.

Roache, Rebecca and Clarke, Steve (2009). 'Bioconservatism, Bioliberalism, and Repugnance', *Monash Bioethics Review* 28(1): 04.1–04.21.

Royzman, Edward (2014). 'Strohminger–McGinn: Deconstructing the Moral Amplification Effect', *Emotion Review* 6(3): 224–5.

Saltzstein, H. D. and Kasachkoff, T. (2004). 'Haidt's Moral Intuitionist Theory: a Psychological and Philosophical Critique', *Review of General Psychology* 8(4): 273–82.

Schnall, S., Haidt, J., Clore, G. L., and Jordan, A. H. (2008). 'Disgust as Embodied Moral Judgment', *Personality and Social Psychology Bulletin* 34: 1096–109.

Tobia, Kevin P. (2015). 'The Effects of Cleanliness and Disgust on Moral Judgment', *Philosophical Psychology* 28(4): 556–68.

Wheatley, Thalia and Haidt, Jonathan (2005). 'Hypnotic Disgust Makes Moral Judgments More Severe', *Psychological Science* 16(10): 780–4.

Woodward, James (2003). *Making Things Happen*. Oxford: Oxford University Press.

Yang, Q., Yan, L., Luo, J., Li, A., Zhang, Y., Tian, X., and Zhang, D. (2013). 'Temporal Dynamics of Disgust and Morality: an Event-Related Potential Study', *PloS one* 8(5): e65094.

Young, Liane and Tsoi, Lily (2013). 'When Mental States Matter, When They Don't, and What That Means for Morality', *Social and Personality Psychology Compass* 7(8): 585–604.

4

Reasons, Reflection, and Repugnance

Doug McConnell and Jeanette Kennett

4.1 Introduction

In a widely cited article, 'The Wisdom of Repugnance' (1997), moral conservative Leon Kass claims that, 'in crucial cases [...] repugnance is the emotional expression of deep wisdom, beyond reason's power fully to articulate it' (1997, p. 20). He argues that 'the burden of moral argument must fall entirely on those who want to declare the widespread repugnances of humankind to be mere timidity or superstition' (1997, p. 21). Kass focuses on repugnance at human cloning and IVF but presumably he would generalize his view to a range of enhancement technologies. He believes that we are too readily 'enchanted and enslaved by the glamour of technology' (Kass 1997, p. 18) and that technology can undermine valuable aspects of our humanity.

In this chapter we draw comparisons between Kass' views on the normative authority of repugnance and social intuitionist accounts of moral judgement which are similarly sceptical about the role of reasoned reflection in moral judgement. We survey the empirical claims made in support of giving moral primacy to intuitions generated by emotions such as repugnance, as well as some common objections. We then examine accounts which integrate intuition and reflection, and argue that plausible accounts of wisdom are in tension with Kass' claim that our inarticulable emotional responses can be the expression of deep wisdom. We conclude that while repugnance and other emotions have a role to play in informing deliberation and judgement, we have reason to be cautious in giving them normative authority. Affective responses alone cannot discharge the burden of justification for moral judgement and are just one tool relied upon by those we consider wise.

4.2 Kass on Repugnance, Moral Judgement, and Wisdom

Kass wrote his article against human cloning in the aftermath of the first successful cloning of a mammal, Dolly the sheep. Despite the public opposition to human cloning,

then as now, he thought that an attempt at human cloning was imminent given a cultural trend to enthusiastically adopt new technology.[1] But even if that were the cultural trend, why should we be concerned about human cloning or any other technology?

Kass complains that certain technologies undermine valuable aspects of our natural humanity. For example, 'cloning shows itself to be a major alteration, indeed, a major violation, of our given nature as embodied, gendered and engendered beings—and of the social relations built on this natural ground. . . . Asexual reproduction, which produces "single-parent" offspring, is a radical departure from the natural human way, confounding all normal understandings of father, mother, sibling, grandparent, etc., and all moral relations tied thereto' (Kass 1997, pp. 20–1). According to Kass, the 'natural human way' is, roughly, that children should be raised by their biological parents who are in a stable monogamous relationship. For Kass, then, there is a fundamental natural humanity which we are morally obligated to protect in the face of technologies that would alienate us from that humanity. Therefore we need to judge which aspects of our humanity need to be protected from which technologies. We do so by feeling repugnance, which generates moral intuitions about the threats to our natural humanity.[2] Speaking of cloning, he says, 'we are repelled . . . not because of the strangeness or novelty of the undertaking, but because we intuit or feel, *immediately and without argument*, the violation of things that we rightfully hold dear' (Kass 1997, p. 20, emphasis added). It is worth noting that for moral repugnance to play this role, it presumably must have a different qualitative character to repugnance at the merely novel. A quantitative difference alone would be insufficient since even the extremely strange presumably falls short of moral wrongness; similarly something morally wrong may, unfortunately, become familiar.

At this point, one might wonder about the direction of explanation here (Giubilini 2015). Must we rely on repugnance to let us know what to hold dear—that is, what counts as our fundamental humanity—or do we define what we hold dear first and then rely on repugnance to alert us if it is threatened? Kass is ambiguous on this point. For the most part he appears to favour the former, affect-driven view, claiming that 'repugnance is the emotional expression of deep wisdom, *beyond reason's power fully to articulate it*' (1997, p. 20, emphasis added). He also argues that the inability to find a justification for repugnance at incest, bestiality, cannibalism, rape, or murder in no way detracts from the normative authority of that repugnance, saying: 'Would anyone's failure to give full rational justification for his or her revulsion at these practices

[1] More specifically, he worries that the growing acceptance of single parenting will make human cloning attractive because clones are 'the ultimate single-parent child'. He also thinks that the modern value that all children should be 'wanted', used to justify abortion and contraception, will be extended to justify cloning people that we 'want'. As it happens, opposition to human cloning *has* decreased, from about 90 per cent to 80 per cent in the last 15 years (Gallup 2015).

[2] Kass is not completely clear on the relationship he envisages between repugnance and intuition. We interpret his view of repugnance as follows. Repugnance is an emotional response which gives rise to an intuitive, contentful judgement that the object of repugnance is morally wrong. The agent might try to justify the intuitions that stem from their repugnance with reflective thought but such justification is optional.

make that revulsion ethically suspect? Not at all' (Kass 1997, p. 20). In other words, we must rely on the normative authority of repugnance to generate intuitions about what we are morally obliged to resist and protect because we often cannot articulate why. Yet he also claims that, 'to pollution and perversion, the fitting response can only be horror and revulsion' (1997, p. 21). This appears to indicate that repugnance should 'fit' with a pre-established idea of what it is for natural humanity to be violated. In another context he mentions a case of repugnance being rightly overruled by reason: 'Surgery involves overcoming repugnance at violating wholeness and submitting to mutilation, overridden here only in order to defend the imperilled body against still greater threats to its integrity' (Kass 2002, p. 183). Kass might also be charitably interpreted as remaining somewhat open to consideration of counterargument when he states that 'the burden of moral argument must fall entirely on those who want to declare the widespread repugnances of humankind to be mere timidity or superstition' (1997, p. 21). There would be no point asking for argument from his opponents if his experience of repugnance was final.[3] Perhaps Kass envisages a kind of mixed model where reason has some normative authority in certain cases (e.g. surgery) while in others repugnance has normative authority (e.g. human cloning). If this is his view he needs to explain how to distinguish the two kinds of case. How do we know when repugnance is wise?

An analysis of wisdom might help us at this point. Strangely, given the article's title, Kass leaves the notion of 'wisdom' completely unanalysed. However, his account resembles Jonathan Haidt's Social Intuitionism, which also gives moral intuition normative authority and may therefore draw on the empirical underpinnings of Haidt's view.

4.3 Affect-Based Accounts of Moral Judgement: Social Intuitionism

Haidt and colleagues (Haidt 2001, 2007, 2012; Haidt and Bjorkland 2008) have proposed that moral judgements are largely the result of moral intuition rather than moral reasoning. Moral intuitions are 'the sudden appearance in consciousness, of a moral judgment, including an affective valence (good–bad, like–dislike) without any conscious awareness of having gone through steps of search, weighing evidence, or inferring a conclusion' (Haidt 2001, p. 818). This account is supported by research that suggests that most, if not all, stimuli activate evaluative associations: an automatic attitude (positive or negative) that is immediate, unintentional, and preconscious (e.g. Fazio 2001). In some cases the evaluative tone of our automatic activations enters awareness in the form of a 'gut feeling'. For example, Haidt (2007, p. 998) notes that '[w]hen we think about sticking a pin into a child's hand…most of us have an automatic intuitive reaction that includes a flash of negative affect.' These automatic evaluative

[3] Though perhaps this request is a mere rhetorical flourish.

intuitions ground or constitute moral judgements which we may then go on to defend or disseminate using more cognitively demanding, slower explicit reasoning. Notably, moral reasoning usually occurs after the automatic production of the moral intuition, and the social intuitionist argues that such reasoning is usually 'a post-hoc process in which we search for evidence to support our initial intuitive reaction' (Haidt 2007, p. 998). This primacy of moral intuition over moral reasoning is allegedly revealed in 'moral dumbfounding' studies in which subjects are presented with harmless but disgusting or taboo-violating scenarios (Haidt and Hersh 2001). Because the negative 'gut feeling' about the act remains—for example, repugnance at the thought of consensual sibling incest—even when the reasons they have cited for condemning it are shown not to apply in the case at hand, subjects continue to insist that the act is morally wrong while acknowledging their inability to explain why.[4]

Based on this kind of evidence, Haidt claims that moral intuitions, rather than reason, cause moral judgements. Haidt (2012) sees reason as a tool to convince people to share one's moral intuitions or form a consensus; consensus creates culturally relative moral truths.[5] Kass, in contrast, believes that his repugnance is alerting him to important and presumably timeless, objective moral truths. Nevertheless they agree in their rejection of reason as a moral guide in key cases, and Haidt appears to share Kass' view that our affectively driven intuitions can be wise, arguing that we should look 'for the roots of human intelligence, rationality, and virtue in what the mind does best: perception, intuition, and other mental operations that are quick, effortless, and generally quite accurate' (Haidt 2001, pp. 821–2). Let us examine the case for intuition before turning to some problems of this approach.

4.4 Wise Intuitions and Unwise Reflection

The folk readily acknowledge cases where acting in accordance with one's emotions is best despite those emotions running counter to deliberative judgement. Nomy Arpaly illustrates one kind of case with Emily, who has always believed that she should pursue a PhD in chemistry. However, she finds that her studies are causing her to feel 'restless, sad, and ill-motivated' (Arpaly 2000, p. 504). Emily sees her feelings as 'senseless and groundless', yet they drive her to leave the programme. At the time of the decision she feels lazy and irrational but also greatly relieved. Only later does she come to understand that her feelings were caused by the fact that the programme was ill-suited to her talents, preferences, and character. In hindsight she 'cites those feelings as the reasons for her quitting, and regards as irrationality not her quitting but that she held on to her conviction that the program was right for her for as long as she did' (Arpaly 2000, p. 504). In a similar case but with a moral flavour, Huckleberry Finn judges it morally

[4] But see Kennett 2011, 2012 for an alternative interpretation of this finding. See also May 2014 for a sceptical view of the influence of disgust on moral judgement.

[5] Haidt endorsed cultural relativism in an interview with Tamla Sommers (2005).

wrong to help the slave, Jim, but continues to do so, motivated by feelings of compassion for him (Bennett 1974). When deliberating he can only find reasons in favour of handing Jim in and none in favour of helping him. Yet he continues to give in to his compassion. As external observers we might see Huck's compassion as the expression of moral reasons that should outweigh those provided by his ideologically clouded judgement. In each of these cases, 'outlaw' emotions appear to function as correctives to internally coherent but false frameworks that structure judgement (Jagger 1996).

How might we explain such cases? How does reflection go so wrong? Conscious, reflective thought is limited. Most obviously, it can be overwhelmed by multiple variables and/or time constraints (Lehrer 2009). Because of these limitations, in some circumstances, intuitive or unconscious cognitive processes can better guide decision-making. An experiment by Dijksterhuis and van Olden (2006), for example, showed that people who carefully reflected on a choice (between five artworks) were more likely to regret their decision than people who were distracted before making their decision. Unconscious thought processes were thought to be superior to conscious ones because,

> first, the unconscious has very high capacity, leading unconscious thought to take into account all information rather than just a subset. Second, unconscious thought…weights the relative importance of different attributes of objects in a relatively objective and "natural" way…Conversely, conscious thought often disturbs this natural weighing process…and can therefore lead to inferior decisions. (Dijksterhuis and van Olden 2006, p. 628)

Similarly, in theories of expert action the focus is on the perceptual and recognitional elements of decision-making. Intuition is often both a faster and a more accurate way to make good decisions than relying on analysis and reflection (Hutton and Klein 1999; Dreyfus 2002; Bortolotti 2011).[6]

This range of evidence raises the possibility that intuitions generated by repugnance could guide wise choices even when they clash with deliberative frameworks that claim to ground good judgement. Repugnance might be the result of unconscious thought processes that integrate a wider range of information than the conscious thought processes of those who argue against that repugnance. Furthermore, attempts to justify repugnance (or a lack of it) may only result in causally irrelevant confabulations that unnecessarily and unhelpfully confound the real situation. However, the picture is more complicated than this evidence might suggest. We will first examine some objections to Kass that raise concerns about the reliability of intuitions, and then review the characteristics of expert decision-making that suggest the intuitions of the wise are not untrained intuitions. We will then turn to an important distinction made by Karen Jones between reason-tracking and reason-responding, before examining folk views of wisdom.

[6] Conscious, reflective thought is still best for some tasks and expert activities—for example, when bookmakers set odds.

4.5 Problems for Kass: Unwise Intuitions

Kass' view faces at least three problems. The first objection targets his particular form of moral realism. The second and third raise doubts as to whether a moral realist can rely on moral intuitions to track moral truth and, therefore, whether intuitions can have normative authority:[7]

1. Kass assumes that he knows what the fundamental aspects of human nature are and that we are morally obliged to protect them. For example, he claims that human cloning is wrong because asexual reproduction is a radical departure from the natural human way and confounds all normal understandings of familial relationships and their moral relations. This raises two problems. First, his view of what is natural for human beings is highly questionable.[8] As Neil Levy points out in response to another moral conservative, Michael Sandel, '[*H*]*omo sapiens* is a niche constructing animal, an animal who shapes the environment that shapes it in turn. . . . We are deeply and essentially cultural animals . . . because it is not just our social organisation, but our intrinsic capacities and traits that are shaped by our culture' (Levy 2009, p. 73). Our niche construction has resulted in our domesticating animals, living in cities, relying on supermarkets, and socializing on the Internet. These cultural and technological changes haven't destroyed our humanity—rather, they are an expression and a development of it. Drawing the line at cloning and other innovations that happen to elicit repugnance in moral conservatives seems arbitrary. There is no reason to think that society is unable to adapt to novel familial relations as the increasing acceptance of same-sex parenting and the evidence that children of such families are as likely to thrive as any others suggests (Crouch et al. 2014). Second, even if Kass was correct about what is natural for humans, he assumes that natural traits are good and worth protecting just because they are natural. However, many things are natural that are not obviously worth protecting (e.g. cancer), so facts about human nature cannot ground normative standards in any straightforward way.
2. Although Kass recognizes that mere novelty may elicit strong negative feelings, he does not explain how the agent can distinguish those cases, where they should discount their initial intuition, from those where they should take repugnance as a veridical moral signal—the expression of deep wisdom. He seems to assume that veridical moral repugnance will have a distinct, unmistakable qualitative character. Such a qualitative character seems unlikely; indeed, evidence from social psychology presented by Haidt and others (e.g. Doris 2006; Prinz 2006) to support their claims that emotion, not reason, is fundamental to

[7] If Haidt is committed to moral truth being relative to one's culture then he avoids these problems.

[8] It's worth noting that this challenge to Kass' definition of moral truth holds whether he arrives at it by repugnance or reason.

moral judgement indicates that our affective responses, and so moral intuitions, can be influenced by morally irrelevant factors. For example, people make harsher moral judgements of people in vignettes if they read them at a filthy desk rather than at a clean desk (Schnall et al. 2008). In another case, participants who were hypnotized to feel disgust when they heard a morally neutral word would judge the behaviour of the protagonist more negatively when that neutral word appeared in the vignette (Wheatley and Haidt 2005). While the results of these experiments should be taken with caution (Jones 2006; Kennett and Fine 2009; Kennett 2012; May 2014) they should concern Kass. If disgust caused by morally irrelevant factors can pollute our moral judgement then its claim to be, or to ground, wisdom is undermined. Disgust does not reliably specify moral targets.

3. Kass seems to accept that some repugnance is misplaced (the merely novel or strange), some repugnance tracks the moral but can be dismissed with good reason (surgery), and some repugnance grounds morality in and of itself (human cloning). But different people make these distinctions in different places: that which elicits a fundamental moral repugnance for one person seems arguable, merely strange, or even familiar to other people.[9] Furthermore, the things people find repugnant change over time. What was once strange becomes familiar. If there is no reason to favour one person's intuitions over those of another then we cannot know whose intuitions are right: repugnance lacks normative authority. Indeed, even moral conservative intuitions have changed over time. For example, conservatives found women who fought for suffrage repellent and unnatural, and their repugnance was widely shared. Viscount Helmsley, speaking in the House of Commons, said:

> It seems to me that this House should remember that if the vote is given to women those who will take the greatest part in politics will not be the quiet, retiring, constitutional women . . . but those very militant women who have brought so much disgrace and discredit upon their sex. . . . One feels that it is not cricket for women to use force. . . . It is little short of nauseating and disgusting to the whole sex. (Hansard Archive 1912)

It is unlikely that latter-day conservatives would regard female suffrage as a departure from human nature or the suffragette's actions as nauseating, however much they might still disapprove of property damage. It is hard to resist the conclusion that much of what conservatives opposed and felt repugnance towards in the past was on account of its strangeness, or its challenge to conservative, white, male privilege, rather than its responsiveness to deep truths about our shared humanity. Perhaps we could settle on the targets that are genuinely repugnant by appealing to the repugnances of the majority, but majority intuitions have clearly been misplaced in the past (e.g. repugnance at

[9] For example, on the basis of his disgust at it, Kass considers licking ice cream in public unacceptable (Kass 1994, p. 148).

interracial marriage). In any case, it is hard to imagine that Kass would be happy to correct his moral judgement on the grounds that the majority didn't share his repugnance. The case of same-sex marriage, for example, is one where conservatives such as Kass currently find themselves in a minority in many societies. Presumably he would see disagreement here as a sign that the majority have lost their moral compass. Therefore Kass needs an argument for why we should believe that *his* repugnances track moral truth; he cannot pass the burden of moral argument to his opponents as he hopes. If the moral judgements that stem from repugnance ultimately require justification, then we have abandoned Kass' view in favour of a more rationalist and deliberative view of moral judgement.

Kass might hope to respond to these objections by appealing to the notion of expertise; repugnance per se does not have moral authority but perhaps the repugnance of moral experts does. The problem for Kass is that research on expert decision-making reveals an important role for slower reflective cognition in the acquisition and deployment of expertise. According to Hutton and Klein, the characteristics of expert decision-making are summarized as:

> (1) Expertise is domain-specific; (2) experts are better able to perceive patterns; (3) experts are relatively fast and virtually error-free; (4) experts have superior memory in their domain of expertise, (although this is often dependent on external cues); (5) experts have a deeper understanding of the problem to solve (e.g. they catch on the causal mechanisms), whereas novices are distracted by superficial features of the problem; (6) experts have a better understanding of their own limitations and an ability to catch themselves when they commit errors; (7) through years of experience, experts acquire the ability to perceive relevant features of the situation (distinguish typical features from exceptions, make fine discriminations, antecedents, and consequences). (Hutton and Klein 1999)

While features (2)–(4) might be automatic, prereflective processes, it is apparent that they are reflectively trained processes rather than completely intuitive ones, and features (5)–(7) require developed reflective abilities, including the ability to conceive of consequences and imagine alternatives.[10] In the moral domain, a deep understanding of a problem would seem to involve the capacity to articulate it and provide a justification for one's response. Consider again the case of Emily. She ultimately concludes that her feelings were a better indicator of what she should do than was her explicit reasoned judgement. It was only much later when Emily achieved a deeper understanding of her ill-suitedness to study—an understanding she could then articulate—that she approached wisdom. For, as the case of the shifting line of moral repugnance towards technological and social change suggests, our feelings, on their own, don't have normative authority. One can easily imagine an alternative ending to Emily's

[10] With regard to (1), Kass could argue that repugnance-based moral expertise is a domain-specific skill or, unlike typical expertise, a relatively domain-general skill. There are issues with both characterizations but we won't pursue them here, given that all expertise involves reflective thought and so doesn't support Kass' view.

story (also amply demonstrated in everyday experience) where she later realizes that her restlessness and lack of motivation in her studies were caused by her internalization of unfortunate gender stereotypes that she now rejects, or by her unacknowledged fears that a successful scientific career would damage her relationship with her mother. She now regrets her earlier decision to quit. In both versions, however, the phenomenology of her coming to quit is exactly the same and it is hard to see why in one case but not the other we should regard the decision as rational or wise *when it was made*. Retrospective endorsement of a decision made on the basis of an intuition is not sufficient.

Nevertheless, we should be cautious about discounting or overriding our intuitions, since, as Karen Jones (2003) argues, they can be reason-tracking. What is needed is a theoretical framework for understanding the relationship between intuitive and reflective cognition, and how both contribute to rational action. We turn now to Jones.

4.6 Reason-Tracking vs. Reason-Responding

Jones begins with the relatively uncontroversial claim that, 'as practical agents, we are trying to latch onto those considerations that really are reason-giving for us in a situation' (Jones 2003, p. 184). The conflict between Kass and those of a more rationalist persuasion is over which mechanisms and methods are best suited to latch on to those reasons—affect and intuition or reflective thought. Based on the range of evidence already considered, it seems that both are essential and each is superior in certain situations. Presumably a wise person would be especially good at knowing when to rely on each. How might this be achieved?

Jones begins to answer this question by making a distinction between reason-trackers and reason-responders. Reason-trackers are capable of 'behaving in accordance with [reasons], but… need possess neither the concept of a reason nor have a self-conception' (Jones 2003, p. 190). A variety of animals count as reason-trackers since innate capacities and learned associations are sufficient to track reasons most of the time. For example, a shark's instincts successfully key it in to what it has most reason to do most of the time, such as what to hunt and when to do it.[11] It is apparent that reason-tracking on its own is not sufficient for wisdom. Sharks may be effective reason-trackers in their domain but they would not, for example, satisfy the conditions for expertise, even within that restricted domain. For that we must also be reason-responders. Reason-responders are sophisticated reason-trackers 'capable of tracking reasons in virtue of responding to them as reasons' (Jones 2003, p. 190). Responding to reasons *as reasons* requires complex reflective capacities in order to see certain reasons as *justifying* one action over alternative actions that may also have something to be said for them. To justify an action, a reason-responder has to be able to step back from their immediate affective responses and question whether they direct us to anything choice-worthy.

[11] Is Jones' definition of 'reason' too broad? Perhaps only reason-responders really have reasons, and beings that are mere reason-trackers do not. This definitional change would not affect our argument.

The advantage of being a reason-responder is that one can track reasons more robustly. 'A creature who lacks critical reflective capacities will not be able to display the same kind of flexibility in its action and sensitivity to the implications of changes in its environment that a reason-responder can' (Jones 2003, p. 90). A shark, for example, will consistently take bait on hooks, unable to recalibrate its reason-tracking mechanisms to distinguish bait from real prey. Most adult humans are reason-responders, which allows them to track their reasons despite living in a highly dynamic sociopolitical world, where reason-tracking strategies need to be regularly revised or replaced. So, for example, one might avoid blood and viscera because one finds it disgusting. Disgust is reason-tracking because it evolved to help us avoid contaminants and disease (Oaten, Stevenson, and Case 2009; Curtis, De Barra, and Aunger 2011). Yet there are specific environments where humans need to overcome that disgust, which lies at the base of repugnance. Trainee surgeons, for example, have to use reflective thought to discount feelings of disgust in order to do what they have more reason to do—save lives and pursue careers. Reason-responders can adjust their behaviour in response to changing circumstances. Intuitive processes are not limited to tracking the reasons they evolved to track, however. They have the potential to reason-track in a range of novel environments but, to do so, they depend on monitoring and recalibration by reflective thought (Jones 2003, p. 196). So, at first, the trainee surgeon needs to try to discount their disgust at each surgical procedure but, over time, their disgust mechanisms adapt so that they are no longer disgusted by surgery.

Despite our self-monitoring efforts we still occasionally find ourselves in situations like that of Emily, wrestling with whether to continue her PhD. Our intuitions conflict with our reflective cognition and we are unsure which is more likely to track our reasons for action. It is tough to know what our reasons are at the time of such conflicts because we have to make do with the self-knowledge our self-monitoring systems have afforded us to date. Such experiences provide opportunities to improve our self-monitoring techniques and self-knowledge so that the next time we face that kind of situation we may better interpret our affective states and question our reasoning more closely.

From Jones' view, then, the rational agent will prefer intuition over reflection when reflective thought is expected to poorly track reasons and/or when intuitions have been trained and so are expected to track reasons accurately. Likewise the rational agent will tend to rely on reason in cases where intuitions aren't suited to the task or where intuitions haven't yet been well trained. The capacity to make such a judgement improves as one's self-knowledge increases, a self-knowledge that depends on the accumulated experience of self-monitoring both one's affective states and one's reflective reasoning. It's worth noting that there is a clear sense in which reflective thought has the ultimate normative authority. Whether one uses reflective thought or intuition at the time of decision, self-monitoring requires reflective thought to assess decisions in hindsight and implement changes for the future. It seems plausible to claim that wise people cannot just be reason-trackers; they must be reason-responders. This

allows them to track reasons despite changes in their environment, their bodies, and their cognitive architecture.

Before considering the implications of Jones' account for Kass, we will say something more about the characteristics of those we consider wise that would allow us to better assess the claim that repugnance can be the expression of deep wisdom. Valerie Tiberius and Jason Swartwood (2011) use the method of wide reflective equilibrium to develop a theory of wisdom that makes no assumptions regarding the truth of any particular moral theory or theory of value. Of particular interest are their conclusions about the epistemic policies that characterize the wise.

4.7 Wisdom: a Closer Analysis

Tiberius and Swartwood developed their view by examining a range of psychological research on folk intuitions about wisdom.[12] The resulting folk theory was then calibrated against theoretical desiderata and research on human cognitive capacities with the goal of arriving at a rationally compelling and psychologically realistic theory of what wisdom consists of. To discover the folk conception of wisdom they conducted a literature review of personal experience and descriptor-rating studies in the psychology literature. In these studies, participants identify and describe wise people and wise acts, or they rate how related and central various characteristics are to wisdom and wise people. The analysis of this literature indicates that there are four key components to the folk concept of wisdom: deep understanding, reflective capacities, problem-solving capacities, and motivation to live well and help others live well.

Wise people have a *deep understanding* of the practical challenges people face, the values that different people hold and how values can change, the ways values impact on practical concerns, and the emotional and intellectual challenges involved in practical success. This includes knowledge of oneself and others (motivations, emotions, habits, skills, limitations, and values), and general knowledge about the physically possible and what can count as a meaningful life. Wise people know 'what matters and can make appropriate distinctions and connections between the various things that matter in life' (Tiberius and Swartwood 2011, p. 282).

Reflective capacities help the wise person develop deep understanding. Wise people assimilate information from diverse sources (personal experience, reflection, and advice from others) and can reorganize old information in new ways. They can apply abstract characterizations of problems to particular states of affairs. Reflective capacities help the wise person 'come to a comprehensive, thoughtful and accurate understanding of the nature of life's challenges and choices' (Tiberius & Swartwood 2011, p. 282).

[12] We focus on the research outlined by Tiberius and Swartwood because we wish to provide an empirically based account of what is required for wisdom in response to Kass' claims—especially insofar as he bases these claims on allegedly shared responses. The theory that emerges from the examination of the empirical literature does chime with much philosophical discussion of wisdom. See, for example, Nozick 1989 and Ryan 1999.

Wise people have good *problem-solving capacities* where they bring their deep understanding to bear on practical problems while remaining mindful of the values at stake (their own and those of others). Practical success requires understanding and responding appropriately to both others' emotions and one's own.

Finally, wise people are kind and helpful; they are motivated to choose well and to help others choose well.

Importantly, the folk take both deep understanding and reflective capacities to be dynamic:

> A wise person is one who guides her actions by her values but also refines and reassesses her values as needed. . . . Most studies emphasise that wise people can take the long-term view of the relationship between their actions, values and priorities, can be self-critical and admit mistakes, and are open to new information that may lead them to change their values. (Tiberius and Swartwood 2011, p. 282)

In other words, understanding can always be made deeper and reflective capacities can always be improved. This is because the world is constantly making new information available and our innate cognitive limitations can be mitigated. The folk therefore see wise people as good reason-responders; wise people self-monitor, train their reflective thought, and build self-understanding. Significantly, the folk also see wisdom as being inherently intersubjective. Wise people learn from others and they are motivated to help other people.

Reflecting on the folk conception of wisdom, Tiberius and Swartwood argue that people need at least three policies to develop and maintain wisdom: policies of *justification*, *epistemic humility and open-mindedness*. These complementary policies can be seen as an elaboration of Jones' view. The wise person creates justificatory stories for how they have chosen to prioritize different values (from multiple perspectives) and distinguish genuine values from things that falsely appear valuable. Justifications can be challenged, developed and taught.[13] By providing justificatory stories, then, the agent opens their decision-making process up to critical help from others. When providing advice, they also give the advisee some basis for discriminating good advice from bad. Epistemic humility is essential once we recognize our epistemic limitations and cognitive biases (biases that influence justificatory reasoning even in quiet moments). If our reasoning can be flawed without our realizing it then we should be prepared to revise our beliefs in the light of new information. Open-mindedness is a good way to access new and potentially challenging information. The wise agent won't automatically revise their beliefs in the face of contrary views, of course, because others also suffer from cognitive limitations. Contrary views signal the need to revisit and compare justificatory stories to see which beliefs and decisions are justified.

[13] Tiberius and Swartwood are alert to the fact that the agent cannot try to justify everything all the time. Rather, reflection can be undertaken in calm, cool moments and then form the basis for more automatic non-reflective decisions when required.

Tiberius and Swartwood claim the view of wisdom they offer is one that people have good reason to aspire to upon reflection, and that they can use to regulate and guide their attempts to develop wisdom and make wise choices (2011, p. 289). By grounding their theory in part in folk views of wisdom they claim to offer 'an ideal that is to some extent already ours' but one which has been refined to take account of our cognitive limitations. It is, they claim, both empirically informed and reflectively sound.

4.8 Wisdom and Repugnance Reconsidered

The view of wisdom gleaned from Jones, and from Tiberius and Swartwood, has significant implications for Kass' claims about the normative authority of deep repugnance.

First, Jones' analysis places clear restrictions on the role repugnance can play in latching on to considerations that count as reasons. To the extent that Kass isolates repugnance from being regulated by reflection or answerable to demands for justification, repugnance will fail to adjust to changes in the environment (or changes in the agent) that might throw it off. We may no longer have reason to avoid some things that innately elicit repugnance, and perhaps we should develop repugnance at things that don't innately elicit it. Repugnance might be reason-tracking in the moral domain, even when it conflicts with reflective thought, but *only* if it has been properly trained by reflective thought (and those reflective capacities themselves have been reflectively trained). Even then we should be cautious in deferring too readily to repugnance responses. The reason for this is that we should expect disgust to produce a lot of false positives—in evolutionary terms we are better safe than sorry when it comes to bad food and communicable disease. Indeed, the evolutionary history of disgust suggests that repugnance may be a particularly *unreliable* guide in social and moral contexts where the fear of contamination too readily attaches to designated out-groups (Kelly 2011). The data on implicit bias should worry us here. For most of us, culturally embedded unconscious associations between, for example, black skin and aggression, or homosexuality and contamination, produce negative affective responses of fear or disgust that drive judgements (particularly when we are tired or distracted) that we would, and should, reject upon reflection, even if we cannot entirely remove the affective responses themselves (Phelps et al. 2000; Dasgupta et al. 2009). On this story the appearance of repugnance should always prompt further reflection to determine whether it is reason-tracking or reason-distracting. The spotted history of moral repugnance does not support the view that even widespread repugnances are likely to be the expression of deep wisdom. Repugnance seems too unreliable to be a feature or expression of wisdom.

Second, the *policies* of the wise identified by Tiberius and Swartwood seem to conflict with the claim that we should trust repugnance, or that the burden of argument must lie with those who do not share Kass' intuitions. A policy of epistemic

humility would encourage the agent to remain open to the possibility that their repugnance was unfounded.[14] Open-mindedness encourages them to consider the views of others who do not share their feelings of repugnance. Finally, the policy of justification requires that any repugnance (or lack thereof) be justified; it cannot settle an argument on its own. Kass might argue that this conception of wisdom is mistaken. Unfortunately he does not offer an alternative account.[15] Moreover, Kass cannot rely on widely shared repugnance to make his argument while rejecting the evidence for a widely shared view of wisdom that tends to undermine his claims about the wisdom of repugnance. There would need to be a good reason to convince the folk that they have a distorted view of wisdom, and that the attributes and policies of the wise should normatively defer to the deep repugnances of moral conservatives.

4.9 Conclusion

Kass (1997) claims moral repugnance can alert us to threats against our fundamental human nature. However, his attempt to define and protect a static human nature is wrong-headed because the evidence supports the view that we are cultural animals whose intrinsic capacities and traits are continually shaped by our culture. Moreover, moral judgements generated by repugnance cannot be trusted to reliably track the moral truth because they are too readily influenced by morally irrelevant factors, and even the repugnances of moral conservatives have changed over time. This all suggests that moral repugnance elicited by new technologies is merely a response to the unfamiliar; at the very least, Kass owes us an argument for why it is not.

Wisdom research suggests that the wise, like other experts, will rely on intuitions over reflective thought but only when those intuitions have been developed through reflection, training, and experience and are subject to reflective oversight. Therefore the normative authority of intuitions is parasitic on long-term reflective training. Repugnance is only one of many ways intuitions are generated and so, even when well trained, repugnance would only be expected to play a minor role in an account of wisdom. More central to wisdom are the policies of epistemic humility, open-mindedness, and a willingness to justify one's actions. These policies allow the wise agent to train both their affective responses and their reflective thinking to track their reasons more robustly. Such self-monitoring is essential because both the world and the agent are constantly changing so that prior training is always going out of date.

[14] Epistemic humility may be consistent with challenging one's opponent to provide reasons why repugnance is not to be trusted, but it seems inconsistent with the claim that those who do not share one's repugnances bear the *entire* burden of proof. Humility would seem to require an acknowledgement that the request for reasons is also one that can legitimately be made of moral conservatives.

[15] We do not deny that some such account might be available, but we do not see it as our task to supply him with an account.

This view of wisdom suggests that we should not rely on our starting intuitions to cast final judgements on novel technologies, including those that promise to enhance human capacities; the very novelty of those technologies entails that our intuitions will not be specifically trained to respond to them (Singer 2005). Wise decisions about how a new technology should be shaped and regulated requires reflective thought and intersubjective reasoning, and, like Emily, we often require more time and experience before we have sufficient knowledge to make our final judgement. In these cases, repugnance may alert us to the need to proceed with caution but it does not have the normative authority to end the conversation.

References

Arpaly, N. (2000). 'On Acting Rationally against One's Best Judgment', *Ethics* 110(3): 488–513.

Bennett, J. (1974). 'The Conscience of Huckleberry Finn', *Philosophy* 49: 123–34.

Bortolotti, L. (2011). 'Does Reflection Lead to Wise Choices?', *Philosophical Explorations* 14(3): 297–313.

Crouch, S., Waters, E., McNair, R., Power, J., and Davis, E. (2014). 'Parent-Reported Measures of Child Health and Wellbeing in Same-Sex Parent Families: a Cross-Sectional Survey', *BMC Public Health* 14(1): 635, <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4082299&tool=pmcentrez&rendertype=abstract>.

Curtis, V., De Barra, M., and Aunger, R. (2011). 'Disgust as an Adaptive System for Disease Avoidance Behaviour', *Philosophical Transactions of the Royal Society B: Biological Sciences* 366: 389–401.

Dasgupta, N., Desteno, D., Williams, L., and Hunsinger, M. (2009). 'Fanning the Flames of Prejudice: the Influence of Specific Incidental Emotions on Implicit Prejudice', *Emotion* 9(4): 585–91.

Dijksterhuis, A. and van Olden, Z. (2006). 'On the Benefits of Thinking Unconsciously: Unconscious Thought can Increase Post-choice Satisfaction', *Journal of Experimental Social Psychology* 42(5): 627–31.

Doris, J. M. (2006). 'Skepticism about Persons', *Philosophical Issues* 19(1): 57–91.

Dreyfus, H. L. (2002). 'Intelligence without Representation: Merleau-Ponty's Critique of Mental Representation: the Relevance of Phenomenology to Scientific Explanation', *Phenomenology and the Cognitive Sciences* 1(4): 367–83.

Fazio, R. H. (2001). 'On the Automatic Activation of Associated Evaluations: an Overview', *Cognition and Emotion* 15: 115–41.

Gallup (2015). 'Cloning', <http://www.gallup.com/poll/6028/cloning.aspx>, accessed 24 August 2015.

Giubilini, A. (2015). 'What in the World is Moral Disgust?', *Australasian Journal of Philosophy* (September): 1–16, <http://www.tandfonline.com/doi/full/10.1080/00048402.2015.1070887>.

Haidt, J. (2001). 'The Emotional Dog and its Rationalist Tail: a Social Intuitionist Approach to Moral Judgment', *Psychological Review* 108(4): 814–34.

Haidt, J. (2007). 'The New Synthesis in Moral Psychology', *Science* 316: 998–1002.

Haidt, J. (2012). *The Righteous Mind: Why Good People are Divided by Politics and Religion*. New York: Pantheon Books.

Haidt, J. and Bjorkland, F. (2008). 'Social Intuitionists Answer Six Questions about Moral Psychology', in W. Sinnott-Armstrong (ed.) *Moral Psychology Vol. 2*. Cambridge, MA: MIT Press, pp. 181–218.

Haidt, J. and Hersh, M. A. (2001). 'Sexual Morality: the Cultures and Emotions of Conservatives and Liberals', *Journal of Applied Social Psychology* 31: 191–221.

Hansard, Archive (1912). 'Debate on the "Conciliation" Bill', *Official Reports 5th Series Parliamentary Debates: Commons*, cols 615–732, <http://www.hansard-archive.parliament.uk/Official_Report,_House_of_Commons_(5th_Series)_Vol_1_(Jan_1909)_to_Vol_1000_(March_1981>, accessed 16 October 2015.

Hutton, R. J. and Klein, G. (1999). 'Expert Decision Making', *Systems Engineering* 2(1): 32–45.

Jagger, A. (1996). 'Love and Knowledge: Emotion in Feminist Epistemology', in A. Garry and M. Pearsall (eds), *Women, Knowledge, and Reality*. New York: Routledge, pp. 166–90.

Jones, K. (2003). 'Emotion, Weakness of Will, and Normative Conception of Agency', *Royal Institute of Philosophy Supplement* 52: 181–200.

Jones, K. (2006). 'Metaethics and Emotions Research: a Response to Prinz', *Philosophical Explorations* 9(1): 45–53.

Kass, L. R. (1994). *The Hungry Soul: Eating and the Perfecting of Our Nature*. New York: Free Press.

Kass, L. R. (1997). 'The Wisdom of Repugnance', *The New Republic* June: 17–26.

Kass, L. R. (2002). *Life, Liberty, and the Defense of Dignity*. San Francisco: Encounter Books.

Kelly, D. (2011). *Yuck!: the Nature and Moral Significance of Disgust*. Cambridge, MA: MIT Press.

Kennett, J. (2011). 'Imagining Reasons', *Southern Journal of Philosophy* 49(Suppl. 1): 181–92.

Kennett, J. (2012). 'Living with One's Choices: Moral Judgment in Vivo', in R. Langdon and C. Mackenzie (eds), *Emotions, Imagination and Moral Reasoning*. New York: Psychology Press, pp. 257–78.

Kennett, J. and Fine, C. (2009). 'Will the Real Moral Judgment Please Stand Up ? The implications of social institutionist models of cognition for meta-ethics and moral psychology', *Ethical Theory and Moral Practice* 12: 77–96.

Lehrer, J. (2009). *How We Decide*. New York: Houghton Mifflin Harcourt.

Levy, N. (2009). 'Neuroethics: Ethics and the Sciences of the Mind', *Philosophy Compass* 4(1): 69–81.

May, J. (2014). 'Does Disgust Influence Moral Judgment?', *Australian Journal of Philosophy* 92(1): 125–41.

Nozick, R. (1989). 'What is Wisdom and Why Do Philosophers Love it So?', in R. Nozick (ed.), *The Examined Life*. New York: Simon and Schuster, pp. 167–278.

Oaten, M., Stevenson, R., and Case, T. (2009). 'Disgust as a Disease-Avoidance Mechanism', *Psychological Bulletin* 135(2): 303.

Phelps, E. A., O'Connor, K., Cunningham, W., Funayama, E., Gatenby, J., Gore, J., and Banaji, M. (2000). 'Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation', *Journal of Cognitive Neuroscience* 12(5): 729–38.

Prinz, J. (2006). 'The Emotional Basis of Moral Judgments', *Philosophical Explorations* 9: 29–43.

Ryan, S. (1999). 'What is Wisdom?', *Philosophical Studies* 93(2): 119–39.

Schnall, S., Haidt, J., Clore, G., and Jordan, A. (2008). 'Disgust as Embodied Moral Judgment', *Personality and Social Psychology Bulletin* 34(8): 1096–109.

Singer, P. (2005). 'Ethics and Intuitions', *Journal of Ethics* 9(3–4): 331–52.

Sommers, T. (2005). 'Jonathan Haidt Interviewed by Tamler Sommers', *The Believer*, <http://www.believermag.com/issues/200508/?read=interview_haidt>, accessed 15 October 2015.

Tiberius, V. and Swartwood, J. (2011). 'Wisdom Revisited: a Case Study in Normative Theorizing', *Philosophical Explorations* 14(3): 277–95.

Wheatley, T. and Haidt, J. (2005). 'Hypnotically Induced Disgust Makes Moral Judgments More Severe', *Psychological Science* 16: 780–4.

5

A Natural Alliance against a Common Foe? Opponents of Enhancement and the Social Model of Disability

Linda Barclay

5.1 Introduction

At first blush it looks as though some opponents of human enhancement and disability advocates share significant normative commitments. Specifically, many from both camps condemn parents or would-be parents who wish to change the genetic or biological features their children would otherwise be born with, a condemnation often expressed by the accusation of eugenics.[1] Despite superficial appearances to the contrary, I will argue in Section 5.2 of this chapter that Michael Sandel's famous argument against meddling with biological givens in the context of enhancement is founded on claims about the value of 'normal' human beings which disability advocates regard as false and pernicious. In Section 5.3 I will assess opposition to changing children's biological givens based on more plausible disability commitments. I will show that these arguments are also weak. Contrary to what disability advocates and opponents of enhancement suggest, it is often a violation rather than an instantiation of parental virtue to abandon judgements about whether one's children's biological givens should be changed.

5.2 Normal and Abnormal

Sandel's well-known opposition to human enhancement centres on his rejection of the kind of hyper-agency, and hyper-parenting in particular, that he believes enhancement

[1] In this chapter I will be concerned only with interventions at the foetal or newborn stage rather than with the issue of selection (e.g. through preimplantation selection, decisions about the timing of conception, etc.).

expresses (Sandel 2004). According to him the drive to mastery displayed by parents' eagerness to enhance their offspring reveals a failure to appreciate the gifted nature of human powers and talents, and to appreciate children as gifts more specifically, accepted unconditionally. The drive to mastery quashes worthy dispositions to be 'open to the unbidden', and thus threatens to destroy our humility and sense of solidarity with one another, as well as to explode the degree to which we are held responsible for the way we are, and for the way our offspring turn out to be.

Sandel does not reject all mastery, does not embrace absolute openness to the unbidden, for, as is well known, he accepts technological interventions aimed at preventing or curing disease. He claims that to

> appreciate children as gifts or blessings is not, of course, to be passive in the face of illness or disease. Medical intervention to cure or prevent illness or restore the injured to health does not desecrate nature but honors it. Healing sickness or injury does not override a child's natural capacities but permits them to flourish. (Sandel 2004, p. 56)

The assumption that there is both a conceptual and a normative distinction between treatment and enhancement is also present in the claim that the moral quandaries associated with enhancement arise 'when people use such therapy not to cure a disease but to reach beyond health, to enhance their physical or cognitive capacities, to lift themselves above the norm' (p. 51).

In asserting that disease is some kind of desecration of nature or derogation of the norm, Sandel assumes that we can demarcate those traits found among human beings into categories of 'natural' or 'normal' on the one hand and 'diseased' or 'abnormal' on the other. He says nothing about disability, but I will assume he intends to assign it to the same category as disease (abnormal, unnatural). It is openness only to the unbidden 'natural' or 'normal' traits that Sandel petitions, not openness to desecrations of nature.

Sandel's opposition to enhancement on the one hand and promotion of treatment for disease and disability on the other is incomplete unless he can demonstrate two claims. First he needs to show that there is some credible way to assign the physical and cognitive traits, talents, and capacities found in human beings into categories of either 'normal' (healthy, non-disabled) or 'abnormal' (diseased, disabled, etc). But more is also needed. Second, even if such categories existed, Sandel would need to muster an additional argument as to why 'natural' or 'normal' traits should be accepted, embraced unconditionally, and why 'unnatural' or 'abnormal' traits are rightly cured or prevented where possible.

With respect to this second claim, numerous of Sandel's critics have already developed a sustained critique of the view that the 'natural' or 'normal' is also and for that reason good (Kamm 2005; Buchanan 2009). The fact that so many opponents of enhancement rarely offer any explicit defence of the assumption that the normal is also good merely invites suspicion that their opposition to enhancement is essentially religious in nature, based on the belief that natural or normal human traits are produced

by an intelligent designer, perhaps in his own image. If one believes, as we should, that natural or normal human traits (assuming such categories exist) have evolved in response to all manner of forces not driven by any design, intention, or conception of good, then the gap between what is natural and what is good hardly requires pointing out.

Disability advocates have a long history of questioning the assumptions that normal traits are good and abnormal one's bad, although their contribution is rarely acknowledged in the enhancement literature. The shared belief of virtually all disability advocates is that the extent to which any given trait, however unusual, limits an individual's abilities or success is either wholly or largely to do with the social environment in which they find themselves, as well as the goals they set for themselves. Here are some representative claims:

> [T]he disadvantage or restriction of activity [is] caused by a contemporary social organisation which takes no or little account of people who have physical impairments and thus excludes them from participation in the mainstream of social activities. . . . Disability is something imposed on top of our impairments by the way we are unnecessarily isolated and excluded from full participation in society. (Oliver 1996, p. 22)

> The dimension and degree of an impairment's interference with an individual's life activity, as well as the line of substantial diminishment of significant performance, are relativised to social arrangements rather than biological states. (Silvers 2000, p. 128)

Some terminological clarification is needed. Many disability advocates accept that some physical, sensory, or cognitive traits of the person are aptly described as 'impaired' (thus perhaps what is meant by abnormal). A person whose eyes can't see has a vision impairment, and a person whose ears cannot hear a hearing impairment. Similarly, people whose legs cannot bear their weight or walk are often described as having an impairment, and so on. But what then is disability? 'Disability' refers to limitations in certain functions, broadly described, such as being able to move from place to place, to communicate, or more broadly still to work, to raise children, or to engage in political activity. As we have seen, disability advocates argue that disabilities are either wholly or at least largely caused by social environments which are incommodious, hostile, or discriminatory against people with impairments. People whose legs are impaired only have difficulty being mobile because of the built environment, and most people with impairments only have trouble communicating or working because of work practices that fail to accommodate the impaired as well as widespread discrimination—hence the moniker 'Social Model of Disability'. In short, despite complex differences in the way disability advocates understand and elaborate upon the Social Model of Disability (Altman 2001), they reject the notion that impaired physical, sensory, or cognitive traits are themselves a significant cause of functional limitation (understood broadly as being able to mobilize, work, communicate, etc.), let alone broader social and political disadvantage. Since from this view abnormal traits are only 'bad' because of incommodious or hostile environments, disability advocates argue that the appropriate

response is to change these environments, not the traits themselves—hence the activist slogan 'Change society, not people'. Importantly, disability advocates have also shown that the assumption that impairments are responsible for functional and social failures has itself been used to justify all manner of discriminatory and other harmful treatment of people with impairments (Silvers 1998). For these reasons, Sandel's opposition to enhancement is built upon assumptions that disability advocates reject, and which they identify as being at the heart of discriminatory treatment of people with impairments—namely, that abnormal or impaired traits are 'desecrations of nature' that should be eliminated.[2]

Some disability advocates also dispute Sandel's first claim—that is, that human traits can be divided into categories of normal or natural on the one hand, and abnormal or unnatural on the other. Ron Amundson argues that human physical, cognitive, and sensory attributes are diverse, and that we have no more reason to corral this diversity into categories of 'normal' and 'impaired' than we have to corral diversity in skin tone, eye shape, and hair texture into racial categories (2000). Philosophers such as Amundson argue that identifying some traits as abnormal or impaired is itself driven by the very stigma and discrimination that disability activists have fought against (Tremain 2001). In opposition, Christopher Boorse has attempted to define disease and impairment in a value-neutral way, based on a notion of deviation from the normal function of a part or process of the human organism (Boorse 1977). If Boorse is wrong (and Amundson and others are right) then Sandel's position is further weakened, not only because he assumes that abnormal traits are bad but because diverse human traits cannot be slotted into categories of normal and abnormal. Whatever the outcome of this debate, it will do nothing to bolster Sandel's other key assumption that normal traits are good and abnormal traits bad. Boorse has always argued that the definition of 'disease' 'neither entails nor is entailed by any therapeutic judgment about people's need for medical treatment' (p. 544). To the contrary, diseases and impairments can be highly desirable under some circumstances: 'Cowpox could save a person's life in the midst of a smallpox epidemic; myopia would be advantageous if it meant avoiding the infantry. Sterility, in a world without contraception, might be a heavenly blessing to parents of large families' (p. 545). Of course, sterility can also be a blessing for parents with small families, or for people with none, which just underlines again how implausible Sandel's

[2] Although disability advocates are right to add their voice to the general philosophical consensus that nothing about the value of any trait follows from its identification as natural or normal, it is not my intention to endorse the stronger view that impaired traits are never in fact sources of disadvantage in hospitable or just social environments. Numerous critics of the social model of disability have rejected the strong view that all functional limitations and broader disadvantage associated with impairments are due entirely to incommodious or unjust environments (Barclay 2010, 2011). Some disability advocates themselves complain that such a perspective ignores the sometimes troublesome and painful features of some impaired bodies, denying a fuller account of the realities of living with impairments (Wendell 1996; Shakespeare 2006). Even if it is true that some impaired traits are truly disadvantageous, even in just social environments, this would not undermine the claim that it is false (and pernicious) to suppose that all impairments or abnormalities should be cured or prevented, which is Sandel's supposition, and our main target here.

conflation of normality with goodness is. Disability scholarship is at its strongest when it lends its voice to the general philosophical consensus about the hopelessness of such conflations.

5.3 Unconditional Parental Acceptance

We have seen that Sandel does not embrace absolute openness to the unbidden, and his reasons are clearly at odds with disability advocacy. Despite this, his claims concerning the appropriate stance of parents towards their children appear to resonate with similar claims made by some disability advocates. He asserts that good parents accept children 'as they come, not as objects of our design or products of our will or instruments of our ambition. Parental love is not contingent on the talents and attributes a child happens to have' (2004, p. 54). Other opponents of enhancement have expressed similar thoughts. Thus Eric Cohen states: 'being a parent means accepting offspring unconditionally' (2006, p. 50). Disability advocates express similar dismay at parents attempting to eliminate impairments. While most of this literature targets selecting against children with impairment through prenatal genetic diagnosis (cf. Asch 1999), there are also a small number of disability advocates who express similar condemnation of parents who would seek to 'cure' their existing children of impairments, either in utero or after birth (cf. Asch and Wasserman 2005).

But why do good parents accept offspring unconditionally, where that is understood to mean not attempting to change the biological or genetic features they are born with? Sandel's explanation in terms of the dangers of 'wilfulness over giftedness, of dominion over reverence, of molding over beholding' (2005, p. 60) has already been successfully critiqued by others (Kamm 2005) so won't be discussed further. Here again we have more to learn from disability advocacy than we do from opponents of enhancement. Indeed, their reasons would seem to be quite obvious given commitment to the Social Model of Disability: if impairments are no more disadvantageous than so-called normal traits in morally acceptable environments, why would parents want to eliminate them unless they falsely and perniciously believe that they are defective?[3]

The problem with the Social Model of Disability is that it apparently also condones parents deliberately *causing* their child to be impaired, or failing to prevent impairment when it is possible to do so (McMahan 2005).[4] Yet even those disability advocates who condemn parents for wanting to eliminate existing impairments do not appear to condone causing impairment or deliberately refraining from preventing it by, for example, withholding a special supplement from a newborn child which would prevent them from developing serious cognitive and physical impairments. For some

[3] I put aside arguments that socially incommodious and hostile environments provide good reasons to eliminate impairments.

[4] In a recent paper Elizabeth Barnes (2014) focuses on the issue of integrity to offer some reasons as to why it might be wrong to cause disability. These arguments have only limited plausibility, however, and would not show why failing to prevent disability was wrong.

critics of the Social Model of Disability this demonstrates that impairment cannot plausibly be evaluated in the way that disability advocates argue—that is, as no more disadvantageous in socially acceptable environments than non-impaired traits.

Elizabeth Barnes is the only philosopher I know of who has developed a serious response to this particular challenge to the disability advocacy position on impairment and disability (Barnes 2009). Although responding to critics of the Social Model of Disability, she does not adopt that model. On the contrary, she aims to offer an alternative to it: while she supports their argument that disability does not necessarily reduce quality of life, she disputes their claim that it is not harmful.[5] Suggesting that disability is just like being gay or being female, Barnes argues that disability is merely another way of being different. Nonetheless, she also argues that disability can be a harm.

What does she mean by saying that disability can be a harm? Disabilities, she concedes, can make life harder—'they present limitations, they cause pain, they subject the bearer to social stigmas and discrimination' (p. 339). However, even if disability is a harm to *local* quality of life, she denies that it follows that it will adversely affect a person's *overall* quality of life. Many things can make life harder—gayness, for example—without necessarily or even reliably making one's life worse off overall. In fact, in the case of gayness, the inference would be deeply offensive. Many people who are gay believe that being gay is a positive thing for them. True, they may experience some harms and disadvantages, but they also experience great benefits—benefits that often outweigh the disadvantages. So despite having a negative impact on some local areas of quality of life, gayness can have positive effects on other local areas such that it can be an overall enhancer of quality of life. So too with disability, according to Barnes. She claims that testimony from disabled people refutes the belief that having an impairment reduces their overall quality of life, even if, on occasion, having an impairment can make life locally harder. She contests the view that gayness and disability are dissimilar in that gayness only makes life (locally) harder because of stigma and discrimination, whereas there are facts intrinsic to the experience of disability that are hardships, in addition to extrinsic features of the social environment. Barnes claims that both gayness and disability present hardships that are a mixture of both intrinsic and extrinsic factors. A gay man cannot have a biological child together with his partner, which seems to be a hardship intrinsic to gayness. Similarly, she suggests, we can concede that some of the hardships of disability are also intrinsic to disability itself even if, as in the case of gayness, most are extrinsic. But whether such hardships are intrinsic or extrinsic, we should reject the inference that local hardships reduce overall quality of life, because testimony does not support such an inference.

What has all this to do with causing a child to have a disability, or failing to prevent a child from acquiring disability? Barnes argues that even if disability does not reduce

[5] Barnes uses the term 'disability' to refer to what I call 'impairment', following as I have standard disability scholarship. For my discussion of Barnes I shall use her terminology of 'disability' to refer to 'impairment'.

overall quality of life, it does not follow that it would be morally acceptable for a parent to cause their child to have a disability, or acceptable for them to refrain from preventing it by, for example, refusing to inoculate the child against polio. For disability is a harm, and, according to Barnes, parents have a duty to prevent serious harms to their children: 'if parents have good reason to believe that something will cause serious harm to their child, they have a duty to prevent it where possible, even if they think that harm might be outweighed by other benefits' (2009, p. 348). Or, as she puts it, parents shouldn't be act consequentialists. Their 'relationship to their children is one based on a duty of care, and a central provision of such a care relationship is the prevention, where possible, of serious harm (regardless of whether such harm might result in eventual benefit)' (p. 348).

While certainly innovative in its defence of standard claims made by disability advocates about the nature and value of being disabled, Barnes' argument is deeply flawed. To begin with, it seems to suggest that parents have a responsibility to prevent their child from becoming gay (if we suppose that it were possible to do so). Barnes has relied heavily on the analogy between gayness and disability, so if parents have a duty to prevent disability, based on the duty not to cause even local harm, then they also have a duty to avoid gayness. But Barnes doesn't accept that parents should prevent their children from becoming gay. She argues that parents are not obligated to avoid all harms but only 'serious' or 'substantial' harms. It is for this reason that they do nothing wrong when they inflict local harms, such as imposing punishments on their children or making them eat their vegetables. These harms, Barnes claims, are 'quite trivial ones' (p. 348). It is only substantial harms—'ones that represent a great deal of suffering'—that parents have a duty to avoid. Barnes suggests that in a society not seriously hostile to gays, being gay would not count as a serious harm. Indeed, she also says that in a society accepting of non-standard physicality which didn't cause 'pain and illness', some disabilities wouldn't count as serious harms either, in which case parents would indeed be under no obligation to avoid them. This concession clearly reins in Barnes' claim to have shown that parents are obligated to prevent, or not cause, disability even though being disabled is just another way of being different.

If Barnes' argument isn't merely another account of the Social Model of Disability, one that bites the bullet on causing disability, she must sustain the claim that many disabilities *do* count as 'serious' harms, even in hospitable environments. Unfortunately she makes no attempt to elucidate the notion of 'serious' harm, and what she says by way of example just raises more questions than it answers. For example, she suggests that polio is a paradigm example of a 'serious disability' and thus one that parents would have a moral duty to prevent. But why? If we can imagine a society that has made substantial advances in accessibility and reducing discrimination, what reasons have we for supposing that the vast majority of physical impairments associated with polio would count as 'serious' harms (or harms at all)? Barnes also appeals to 'pain and illness' as features that can make a disability a serious harm. However, few disabilities are typically associated with chronic, ongoing pain—certainly not blindness, deafness,

severe cognitive impairment, and many forms of paraplegia or quadriplegia. In the absence of any definition of illness, or of how it is distinct from disability, and in the absence of any argument as to why we should believe illness is a serious harm whereas disability is not, the appeal to 'illness' is hardly helpful either.[6] The upshot is that it is unclear whether Barnes has really provided any rejoinder to people such as Jeff McMahan who challenge widespread disability advocacy views for their implausible implications regarding causing or failing to prevent disability. Given that she relies so heavily on the analogy between disability and gayness, her rejoinder fails without some account of why disability, but not gayness, is a serious harm.

However, let's assume for the sake of argument that disability is sometimes a serious harm. There are still significant problems with Barnes' view. It is implausible to suggest, as she does, that good parents do not inflict local harms even in the pursuit of enhancing overall quality of life. Imagine my son gets cancer. I am told that with chemotherapy he has a very good chance of being cured. The chemotherapy will cause all kinds of harm: he will suffer weakness and nausea, serious skin irritations, hair loss, and perhaps even developmental delays. Would it really be ideal parenting to decline the treatment regime on the basis that I should avoid the infliction of serious local harms even if in the pursuit of my child's overall benefit? It won't work for Barnes to emphasize that the child will die if the harm is not inflicted, for that too is a judgement about the child's *overall* outcomes, not merely one about whether the infliction of the harms of treatment are 'serious' or 'substantial'. In any case, we can simply change the details of the case by supposing that the cancer will not kill my son but just mean that he will live a life of much reduced physical strength, eventually leading to wheelchair use, blindness and deafness, and eventual loss of speech. In at least many cases, we do and should impose local harms on children where we have good reason to believe that doing so will lead to significant overall benefit for them. It is incredible to suggest that good parents should always avoid making such judgements. Indeed, imposing local harms in the name of promoting *overall* quality of life outcomes is the common grist of good parenting. We do not feel entitled to impose discipline on our children or make them eat their vegetables merely because we are sure that the harms involved are 'quite trivial ones', as Barnes suggests. We do so because of our beliefs about what is likely to be good for their overall quality of life. Indeed, I suspect that whether or not we classify a harm as 'serious' or 'trivial' in many cases will depend on whether it contributes to or detracts from overall quality of life. Making children eat broccoli is a trivial harm mainly because it actually contributes to their overall quality of life. Making them eat some harmless but foul-tasting substance for no good benefit at all does seem like a more serious harm. Similarly, the imposition of good-quality discipline is a trivial harm because it is good for children overall; the imposition of pointless and capricious

[6] In her latest paper (2014), Barnes also relies on a distinction between health and disability to avoid unpalatable conclusions, but there too she fails to explain what she takes the distinction to be or why she thinks it has normative significance.

discipline which in no way improves a child's quality of life could plausibly be described as a serious harm.

Good parents will want to know of any disability not only whether or not it will be a local harm but what its impact on overall quality of life might be. Because Barnes denies that disability causes an overall reduction in quality of life, she ultimately fails to explain why parents should not cause, or fail to prevent, disability. But things might be even worse for Barnes. She argues that the testimony of disabled people supports the view that being disabled can in fact enhance overall quality of life. Once we accept that parents should make overall quality-of-life judgements, and in some cases appropriately inflict harms in the pursuit of enhanced quality of life, it is unclear why parents who refuse to cause disability don't do something wrong, at least in some cases. Barnes suggests an alternative, more modest view of the connection between disability and quality of life that might evade this challenge: not the strong claim that disability enhances quality of life but merely that we cannot know whether it will, or whether it will diminish it instead. She claims that what we want to know of any individual is their objective chance of a good quality of life with a disability, and that to determine this

> it's not enough to simply determine the average quality of life of persons in similar circumstances with similar disability and compare it to persons in similar circumstances without disability. Such a calculation (were it possible) would only tell us about average quality of life. But what we need to know is whether disability is objectively likely to be a determent for x. (2009, p. 343)[7]

By parity of reasoning, Barnes must believe that parents cannot make informed calculations as to whether a healthy diet and good education are objectively likely to enhance their children's quality of life. Even if she is prepared to bite that bullet, one final question remains. Why does Barnes presuppose that we *can* make an assessment of whether or not disability is a serious local harm? Presumably the best we can do is consider the evidence, including the testimony of disabled people themselves, which is just as varied regarding the issue of disability as a local harm as it is regarding the question of whether disability raises or reduces overall quality of life.

The upshot of the discussion is as follows. If having an impairment in a hospitable environment is of no disadvantage, then disability advocates can lend some credibility to the criticism of parents who would seek to change their children's biological givens. Specifically, they have the resources to flesh out the slur of eugenics by suggesting that parents who would seek to change their children's biology are acting on nothing more than false and pernicious views about the value of some human traits and capacities. They provide some substantive support for being 'open to the unbidden'. However, the apparent problem with this argument is that it would seem to entail that there is nothing wrong with parents causing their children to be impaired. Barnes fails to provide

[7] Barnes does go on to concede that perhaps such judgements are possible, but then claims we would have no way of knowing whether reductions in overall quality of life were due to the intrinsic features of the impairment or social circumstances. I am here assuming that we are discussing cases located in a hospitable social environment.

a convincing response to this challenge. Given that so few, if any, disability advocates are willing to bite the bullet and concede that it is morally acceptable to cause impairment,[8] then ultimately they haven't presented us with a compelling reason why parents should accept their children as they are. Their position is more credible than Sandel's vague claims about the importance of accepting and unconditional love, but in the end it also fails.

5.4 Conclusion

There is much to be learned from disability scholarship but I don't think we learn that parents should never seek to prevent or cure impairments. We do learn the more modest but important lesson that it is extremely difficult to make certain judgements about what will or will not enhance quality of life. We have very little reliable evidence as to what quality of life would be like for people with impairments in a social environment that was commodious and non-discriminatory. We have no experience of such an environment, and near-universal prejudice against impairment and people with impairments means our own speculative assessments are likely to be heavily tarnished. Even in the absence of such limitations, it is empirically difficult to assess quality of life. There is ongoing controversy about how to interpret available evidence, and about the quality of the evidence available. The empirical difficulties are only compounded by philosophical disputes about what quality of life consists of. Perhaps it is safe to say that severe mental illness, pain, chronic loneliness, and social hostility are certainly not part of it, and that contentment certainly is. Beyond that, things become much less certain. When people appeal to evidence that people with impairments are just as happy with their lot as people without impairments, they assume that quality of life is concerned only with subjective well-being or happiness, and that the best evidence really does show that people with impairments are as subjectively content as those without them. Both assumptions are contestable. Similarly, the pro-enhancement literature is riddled with equally contentious assertions about how greater tallness, intelligence, and memory increase overall well-being, engaging in cherry-picking of limited literature and displaying questionable confidence in what that limited literature can really demonstrate (cf. Savulescu, Sandberg, and Kahane 2011).

Disability scholars are right to claim that we should not assume that all, or even most, impairments are intrinsic disadvantages, irrespective of social circumstances. Our fixed traits can interact with the social environment in all kinds of ways to produce advantage and disadvantage (variously described, depending on one's philosophical view about well-being). So too with respect to the kinds of trait that may in the future be amenable to enhancement technology. Because such judgements in most cases are both empirically and philosophically uncertain, we should not assume that either curing disability

[8] This may explain why, in her most recent paper, Barnes simply disputes the claim that there is anything wrong with causing impairment (2014).

or enhancing normal traits is always a good thing. Despite all the uncertainty involved, at best we reach a conclusion that parents and others should proceed with great caution. Sandel is right to promote the virtue of humility with respect to such matters. However, being humble and cautious about our ability to form reliable judgements on quality of life is hardly a broadside against human enhancement, any more than it is a general criticism of parents who seek to prevent or cure disability. Even the most enthusiastic advocates of human enhancement agree that we should proceed with caution, and acknowledge that our judgements may be tainted by prejudice, fads, and ignorance. Perhaps we should give up our obsessive focus on trying to cure some impairments, as well as our misplaced optimism that some enhancements will do anyone any real good. What we cannot conclude is that we should give up trying to make such judgements altogether, that we should just be open to any and all of the unbiddens our children will otherwise turn up with. In any case, such a posture would not reflect virtuous parenting but in many circumstances an abandonment of it.

References

Altman, Barbara M. (2001). 'Disability Definitions, Models, Classification Schemes, and Applications', in Gary L. Albrecht, Katherine Delores Seelman, and Michael Bury (eds), *Handbook of Disability Studies*. Thousand Oaks, CA: Sage Publications.

Amundson, Ron (2000). 'Against Normal Function', *Studies in History and Philosophy of Science* 31(1): 33–53.

Asch, Adrienne (1999). 'Prenatal Diagnosis and Selective Abortion: a Challenge to Practice and Policy', *American Journal of Public Health* 89(11): 1649–57.

Asch, Adrienne and Wasserman, David (2005). 'Where is the Sin in Synecdoche? Prenatal Testing and the Parent-Child Relationship', in David Wasserman, Jerome Bickenbach, and Wachbroit, Robert (eds), *Quality of Life and Human Difference: Genetic Testing, Health Care, and Disability*, New York: Cambridge University Press, pp. 172–216.

Barclay, Linda (2010). 'Disability, Respect and Justice', *Journal of Applied Philosophy* 27: 154–71.

Barclay, Linda (2011). 'Justice and Disability: What Kind of Theorizing Is Needed?', *Journal of Social Philosophy* 42(3): 273–87.

Barnes, Elizabeth (2009). 'Disability, Minority, and Difference', *Journal of Applied Philosophy* 26(4): 337–55.

Barnes, Elizabeth (2014). 'Valuing Disability, Causing Disability', *Ethics* 125(1): 88–113.

Boorse, Christopher (1977). 'Health as a Theoretical Concept', *Philosophy of Science* 44: 542–73.

Buchanan, Allen (2009). 'Human Nature and Enhancement', *Bioethics* 23(3): 141–50.

Cohen, Eric (2006). 'Conservative Bioethics and the Search for Wisdom', *Hastings Center Report* 36(1): 44–56.

Kamm, Frances M. (2005). 'Is There a Problem with Enhancement?', *The American Journal of Bioethics* 5(3): 5–14.

McMahan, Jeff (2005). 'Causing People to Exist and Causing People to be Disabled', *Ethics* 116(1): 77–99.

Oliver, Michael (1996). *Understanding Disability. From Theory to Practice*. Basingstoke: Palgrave Macmillan.

Sandel, Michael J. (2004). 'The Case Against Perfection', *The Atlantic* April: 51–62.

Savulescu, Julian, Sandberg, Anders, and Kahane, Guy (2011). 'Well-Being and Enhancement', in Julian Savulescu, Anders Sandberg and Guy Kahane (eds), *Enhancing Human Capacities*. Oxford: Wiley-Blackwell, pp. 3–18.

Shakespeare, Tom (2006). *Disability Rights and Wrongs*. Abingdon: Routledge.

Silvers, Anita (1998). 'A Fatal Attraction to Normalising: Treating Disabilities as Deviations from "Species-Typical" Functioning', in Erik Parens (ed.), *Enhancing Human Traits: Ethical and Social Implications*. Washington DC: Georgetown University Press, pp. 95–123.

Silvers, Anita (2000). 'The Unprotected. Constructing Disability in the Context of Antidiscrimination Law', in Leslie Pickering Francis and Anita Silvers, *Americans with Disabilities. Exploring Implications of the Law for Individuals and Institutions*. New York: Routledge, pp. 126–45.

Tremain, Shelley (2001). 'On the Government of Disability', *Social Theory and Practice* 27(4): 617–36.

Wendell, Susan (1996). *The Rejected Body. Feminist Philosophical Reflections on Disability*. New York: Routledge.

6

Playing God

What is the Problem?

John Weckert

6.1 Introduction

To be accused of playing God is to be accused of doing something morally wrong, an accusation that is often levelled at human enhancement (see Peters 2007). The wrongness might derive from harms caused but it is not always seen in this light. It might rather come from the fact that something is being done that humans have no business doing. Speaking of eugenics and genetic engineering, Michael Sandel asks: 'Why should we not shake off our unease about genetic enhancement as so much superstition?' His answer is: 'To believe that our talents and powers are wholly our own is to misunderstand our place in creation, to confuse our role with God's' (Sandel 2004). Clive Hamilton makes a similar point regarding geoengineering:

> Playing God entails humans crossing a boundary to a domain of control or causation that is beyond their rightful place. (Hamilton, 2013, p. 178)

And again:

> To cross successfully would require mortals to possess a degree of omniscience and omnipotence that has always been preserved for God or the great processes of Nature that are rightfully beyond human interference. (Hamilton 2013, p. 180)

So we play God when we are 'beyond our rightful place' and in a place that is 'rightfully beyond human interference'. There is a hint in the second quotation that playing God will cause harm because humans do not have the capacity to make good decisions in those areas, but the wrongness essentially consists in interfering where we should not. Henk van den Belt says this explicitly when discussing Frankenstein—'his real transgression was that he sought to create a *human* being'—implying that the real transgression was not that his action led to harm (van den Belt 2009, p. 267).

The creation of a monster by Frankenstein (Shelley 1818) can be seen as a paradigm case of playing God. The most common use of this image currently is in

the term 'Frankenfoods' to designate foods that have been genetically modified (GM) by scientists. The modification of genes, the assumption goes, is the province of God and not of humans. The main concern of many, however, seems to be the real or feared harms caused by producing and consuming those foods, so perhaps it should be seen as a largely rhetorical device to discourage the consumption and production of GM crops. First used in 1992 by Paul Lewis (Medicinenet 2012), the term 'Frankenfood' became popular after the UK's Prince Charles, the Prince of Wales, strongly criticized GM crops in the late 1990s and linked them with playing God:

> I believe that genetic modification is much more than just an extension of selective breeding techniques. Mixing genetic material from species that cannot breed naturally, takes us into areas that should be left to God. We should not be meddling with the building blocks of life in this way. (Prince of Wales 1998)

It appears from this as if the wrongness consists in doing something that is rightfully in God's domain, but the Prince then talks primarily about 'the long-term consequences for human health and the environment' of these crops. The wrongness really seems to lie essentially in the potentially harmful consequences.

Playing God has a distinctively religious ring. However, van den Belt (2009) argues that liberal theologians do not see it as primarily a religious argument but rather a secular one. It is that humans infringe not on God's domain but on a deified nature's domain. For Ted Peters, one such theologian, playing God has its origins in Greek religion and explicitly in the Prometheus myth, where by taking fire from the sun and giving it to humans, Prometheus moved into an area that was rightfully Zeus' (Peters 2007, p. 179). According to Peters but probably not to all Christians, in the Christian tradition the playing God argument has no place. New technologies are not to be feared because they might infringe on God's territory. What is important is whether these technologies 'will respond to a transcendental ground for goodness and will enhance our capacity to love' (Peters 2007, p. 182).

Whether or not there is a plausible religious version of the playing God argument will not concern us here because the secular version is now more common. In this version, as already noted, nature plays the role of God. Instead of claiming that some things are in the province of God and not humans, here some things should be left to nature. If humans attempt to do what is rightly in nature's domain, it is in some sense unnatural.

The claim then is that playing God is interfering with nature or modifying nature, as in the genetic engineering example, and that this is morally wrong. The argument to examine these claims will proceed as follows: if playing God is interfering with nature, some account of nature is required. It will be argued in Section 6.2 that there are two broad concepts of nature, one excluding and the other including humans. Section 6.3 will consider what follows for the playing God argument if humans are not part of nature, and Section 6.4 the consequences if they are. Neither case gives strong support

to the claim that playing God as interfering with nature is morally problematic. Sections 6.3 and 6.4 suggest that no useful distinction can be drawn between the natural and the artificial that raises problems for interfering with nature, but in Section 6.5 an alternative distinction—that between what is controlled by humans and what is not—is explored, and this forms the basis for the rest of the argument. In Section 6.6, playing God is spelt out in terms of extending the sphere over which humans have control. Examples are used to demonstrate this in Section 6.7, and finally, in Section 6.8, some conclusions are drawn.

6.2 What is Nature?

The focus here is on nature rather than on what is natural. The starting point, but only the starting point, is that nature is what I see when I'm in the country: trees, grass, animals, and so on. This can, of course, be described in terms of what is natural, but a focus on nature makes clearer the relationship between the secular version of playing God and the religious. Nothing of substance hinges on this distinction between nature and the natural.

Moving on to a slightly more precise account, from our starting point of nature being what we see when we are out in the country, a common view is that nature is what we are in when we go out into the mountain or forests. That is, nature is where people are not. Or perhaps better, nature is what has not been affected by people. If this is what nature is, there is not much of it around. These days it might be more usual to think of wilderness as that which has not been affected by humans, but this too is dubious. According to the Wilderness Society (TWS) in Australia, one criterion of wilderness is to be 'substantially undisturbed by colonial and modern technological society'. A distinction is drawn 'between the impacts of indigenous people practising ecologically sustainable land management, and the very different impacts of colonial and modern technological society.' (TWS 2015). Even wilderness is recognized as being affected by humans.

Before considering this issue of whether something can be nature or wilderness and simultaneously affected by humans more closely, another question arises: are people part of nature? Fern Wickson raises this well:

> If nature is somewhere that humans are not, we lose sight of the fact that we are just another species intimately intertwined in the complex web of biological systems on this planet. However, if we place ourselves within a definition of nature, the definition then becomes essentially meaningless by extending to everything on Earth. (Wickson 2008)

Either nature does not include humans or it does. The former is wrong, she suggests, but the latter makes the concept 'nature' meaningless because everything is part of it.

This is only an apparent dilemma. First, if humans are part of nature, the concept is not necessarily meaningless if it is interpreted as meaning that humans are part of the natural world, just as kangaroos and polar bears are. Second, the argument here is that

there are two broad conceptions of nature: one that includes humans and the other that excludes them. John Stuart Mill suggested this:

> It thus appears that we must recognise at least two principal meanings in the word 'nature'. In one sense, it means all the powers existing in either the outer or the inner world and everything which takes place by means of those powers. In another sense, it means, not everything which happens, but only, what takes place without agency, or without the voluntary and intentional agency, of man. (Mill 1904)

John Passmore too makes the distinction when he writes that he is using 'nature' in a narrow sense,

> in which neither Sir Christopher Wren nor St Paul's Cathedral forms part of 'nature' and it may be hard to decide whether an oddly shaped flint or a landscape where the trees are evenly spaced is or is not 'natural'. (Passmore 2010, p. 103)

Distinguishing these two broad conceptions of nature is important for the following discussion, although it will be argued that the distinction is not a sharp one and that it is actually more of a continuum, with artefacts at one end and something like wilderness at the other, where the two broad concepts of nature are the limiting cases. But more of that later. Even if the distinction is not sharp, it is still useful. Call the concept of nature which includes humans as well as the non-human, nature(I) and that which excludes humans nature(E). Given this distinction, a beaver's dam is both part of nature(I) and nature(E), while a rock-and-concrete dam built for irrigation is part of nature(I) only.

What implications does the discussion so far have for the playing God argument? We will consider the implications of the two concepts separately.

6.3 Humans Are Not Part of Nature

In the religious version of playing God, we should not do some things because they are in God's domain, not ours. In the secular version, if nature is nature(E) then we should not do certain things because they are in the domain of nature(E); some things should be left to nature(E). Humans should not interfere with nature(E). While this secular version seems to run closely parallel to the religious one, at least at the start, a puzzling question is raised: why should humans not interfere with nature(E)? This question can, of course, also be asked about moving into God's domain, but there the answer might be sought from theologians or revelation. It is less clear who has the authority to tell us why nature should not be interfered with. Perhaps we should not do anything that is not natural, but that seems to just mean that we should not do anything that interferes with nature(E)—hardly an advance. An argument is required to show why we should not interfere with nature(E)—that is, do things that are not natural. An argument can be made that interfering with nature is natural for humans. Natural actions must include those that are necessary for survival—eating, for example. Relative to other

animals, humans are not very good at many things. We cannot run very fast, certainly not fast enough to catch most of our prey; we swim only poorly; we cannot fly at all; we clumsily climb trees; and we are not adept at burrowing. And our senses seem to be middle range at best. Where we do excel is in our reasoning and in our tool-making ability. By and large we use our tool-making ability, our technologies, to provide our food and shelter, and in general to enable us to lead reasonable lives. According to Jose Ortega y Gassett (1961), the purpose of technology is to enable us to live well. Without those tools we would not survive and certainly not live comfortably or well, but tools and their use interfere with nature(E). Even indigenous Australians who seemed to live in relative harmony with nature(E) used fire to modify the landscape in order to increase their food supply (Jones 1969), and the world over, land has been cleared of its vegetation so that food crops can be grown. There is something very natural about humans interfering with nature(E): we are just doing what we evolved to do. There is no choice if we are to survive. It follows from this that if humans are not part of nature, the argument that we should not interfere with it does not take us very far. Playing God cannot be simply interfering with or modifying nature(E); something else is needed. Perhaps it is the way or the extent to which nature is modified, but we will return to this in Section 6.6.

6.4 Humans Are Part of Nature

Here it is even more difficult to get the argument going that humans should not interfere with or modify nature. If nature is nature(I) then whatever humans do should be viewed in the same way as what non-humans do. Birds build nests while humans build houses. Beavers build dams out of trees and sticks, while humans build dams out of rocks and concrete. Houses and concrete dams are just as much part of nature(I) as nests and beaver dams. To claim that we should not interfere with nature(I) makes little sense.

Perhaps it could be argued that while humans are part of nature(I), we are different in kind and not just in degree from the rest of nature(I). We are autonomous in that we have free will and no other creatures do. We are therefore responsible for what we do but other creatures have no responsibility. One response to this is to give a compatibilist account of free will in which everything that we do is caused, just as everything that other creatures do is caused, albeit in a much more complicated and sophisticated manner. On such an account it would be wrong to say that we are different in kind; we would differ in degree only. Another response is to accept that we are different in kind and not merely in degree. That is just the way that we evolved. Kangaroos evolved with strong tails to jump and we evolved with a decision-making capacity that makes us autonomous. It is not obvious why autonomous creatures could not be part of nature, and if they are, what they do is part of nature. The claim that we should not interfere with nature because only we are autonomous is difficult to justify. Again, it might be that what is important is how it is modified.

So far it looks as if the secular version of the playing God argument gets nowhere, but the story of Frankenstein can be instructive, as will be seen in Section 6.7. First, though, a distinction will be spelt out that will prove useful. Given that it has been argued that whether or not humans are considered part of nature, nothing of significance can be established to show that interfering with nature is morally dubious, so any distinction between the artificial and the non-artificial is not likely to be useful. Another distinction—that between what is and what is not controlled by humans—is promising.

6.5 The Controlled/Uncontrolled Distinction

If humans are part of nature then everything that they do is natural, and if they are not part of nature, much of what they do is still natural because they have no choice but to interfere with nature if they are to survive. They do what they have evolved to do. They modify nature just as beavers and nest-building birds do. So for now we will put aside the question of whether what humans do is natural or not and consider instead the controlled/uncontrolled distinction. Something or some aspect of the world is controlled if it is made by humans (on the assumption that they can control it), and if it is not then it is uncontrolled. A concrete-and-rock dam is controlled while a beaver's dam is uncontrolled (by humans). Landscapes intentionally modified by indigenous Australians through intentional and systematic burning are controlled while wilderness areas untouched by humans, if any remain, are uncontrolled. Another way of putting this distinction is in terms of artefacts and natural objects. The concrete-and-rock dam is an example of the former and the beaver dam one of the latter. This distinction as stated is ambiguous. Something could be uncontrolled by humans either because nobody has yet bothered to try to control it or because it is uncontrollable, because either the technology is not yet available to control it or because it is uncontrollable in principle. Here 'uncontrolled' will mean that nobody has yet either controlled it or developed the technology to control it—that is, it embraces everything except uncontrollable in principle.

It will be obvious that this distinction is not a sharp one. Something can be controlled to a greater or lesser extent, and given that this distinction can be put in terms of artefacts, being an artefact can also be a matter of degree. Lynne Rudder Baker (2008, p. 7) makes a stronger claim, arguing that the artefact/natural object distinction will continue to become fuzzier and will eventually fade away. This is at least partly because it is unclear if 'digital organisms, robo-rats, bacterial batteries, genetically engineered viral search-and-destroy missiles' are artefacts or natural objects, nor is it obvious whether it matters. Val Plumwood (2006) too suggests that the distinction between artefacts and natural objects is not sharp. Her main interest is the landscape, and she argues that any landscape is a result of not only human agency but also the agency of other species and of nature itself. Because of this, she argues against Tim Flannery that Australia is

not a vast human artefact, even if the landscape was shaped to a large extent by intentional burning (see Flannery 2003, p. 41). If agents interact in the creation of artefacts, then it seems to follow that the greater the human involvement, the greater degree that an object is an artefact. A farm is an artefact to a greater extent than a wilderness area even if the current state of the wildness has been affected to some degree by human activity. This point has been laboured a little here because it is important in what follows.

It should be noted that this distinction between controlled and uncontrolled is species relative. From a human perspective, houses are controlled and birds' nests are uncontrolled. From the perspective of birds, if they have one, it is the other way round. For them a bird's nest would be an artefact but my house part of nature. C. S. Lewis also makes this point:

> If ants had a language they would, no doubt, call their anthill an artifact and describe the brick wall in its neighborhood as a *natural* object. (Lewis 1960, pp. 45–6)

This may seem a little odd, but nothing important hinges on this oddness.

6.6 Back to Playing God

The idea that interfering with or modifying nature is illegitimate because it is unnatural leads nowhere, as we have seen. If nature includes humans then everything that we do is natural. If it does not include humans then an argument is still required to show that interfering with nature is somehow wrong. As was argued, humans in fact have no choice but to interfere with nature in order to survive, and if there is no choice then it can hardly be wrong. It might be argued that only interference necessary for survival is justifiable, but this places a lot of weight on 'survival'. If survival is just staying alive long enough to reproduce then perhaps not much interference would be necessary, but to lead a satisfying human life more than mere survival is necessary, so more modification of nature is required. How much more is justified needs to be spelt out, but that is beyond our scope here. We are now, however, heading in the right direction.

We will return now to playing God with the controlled/uncontrolled distinction in mind and see whether that is more fruitful than looking at what is unnatural. Playing God, we might say, is expanding the sphere over which we have control. This is obviously much too broad as it stands. Every time we clear a small patch of native forest to plant a crop for food we would be playing God, but that would make the whole idea insignificant. One way to make this broadening of sphere more significant is to consider what Hans Jonas says about ethics and technology:

> All previous ethics [...] had these interconnected tacit premises in common: that the human condition, determined by the nature of man and the nature of things, was given once for all; that the human good on that basis was readily determinable; and that the range of human action and therefore responsibility was narrowly circumscribed. (Jonas 1973, p. 31)

The suggestion is that in the past, technology was not powerful enough to change either humans or the non-human world, so our decisions were made and therefore our responsibilities restricted within fairly clearly defined parameters. If we clear our small patch of land by hand or with some hand-held implement, say a hoe, we are not able to make significant changes to 'the nature of things'. Everything, apart from a small patch, is much the same as it was before the clearing was done. Jonas goes on to argue that new technologies have changed all this and now ethics, and our responsibilities, have to be reassessed in the light of these new situations. The range of human action and responsibility is no longer so narrowly circumscribed.

Ronald Dworkin makes a similar point:

> That crucial boundary between choice and chance is the spine of our ethics and our morality, and any serious shift in that boundary is seriously dislocating. (Dworkin 2000, p. 444)

In the area over which we have control, we have choice. In the rest it's up to chance. Again the suggestion is that in the past this boundary did not shifted dramatically but now it has.

These quotations point to an understanding of playing God which does give it some important content, or so it will be argued. Playing God, we will say now, is expanding the sphere over which we have control to the extent that we no longer have the knowledge or capacity to competently make decisions with respect to the expanded part. C. A. J. Coady compares the ability of humans with those of the Christian God, who is considered to be 'omnipotent, omniscient and supremely benevolent' (Coady 2009, p. 163). Humans are none of these and can therefore easily move into areas where they lack competence.

To make this clearer, consider an example. The Wivenhoe Dam was built on the Brisbane River partly for flood mitigation. In the past the river had caused severe flooding in the city of Brisbane and it was hoped that the dam would alleviate the situation. In 2011, torrential rain caused major flooding in Brisbane again. However, now that the dam had been built, the engineers in charge were held by many to be responsible for the flooding. If water had been released earlier, it was claimed, the severity of the flooding would have been considerably less. It is not that the engineers in charge of water release were incompetent. Neither they nor anyone else had the necessary knowledge. Torrential rain could be predicted but just how much and where it would fall could not be known accurately. Furthermore, the area had been in drought for a considerable time, so releasing water if the rain did not eventuate could have led to future water shortages. Here we have a situation where an event—the flooding—which would previously have been an act of God, or a chance event, is now the responsibility of people. In Dworkin's terms, the boundary between choice and chance had changed and in Jonas' terms, human responsibility is no longer so narrowly circumscribed. The claim here is not that building the Wivenhoe Dam was playing God. Rather, it is an example of how boundaries and responsibilities can change, and how decisions must be made in situations where previously none were necessary.

It should be noted here that we are not talking about all, or even most, situations in which decisions must be made where they were not previously. If I buy a car with manual transmission when previously I have always had cars with automatic transmission, I have new decisions to make. I have to decide when to change gears when previously I did not need to. The car took care of itself in that respect. This differs from the Wivenhoe Dam example. In that case, what had been a chance event—a flood—was now under the control of humans. The automatic transmission was not a chance event in this way. The gears changed as a result of the design of the transmission. In effect, the gear change was now an effect of my decision where previously it was the result of someone else's—the designer's—decision.

The situations of interest are those where previously no human decision-making was possible but now it is. Even here, however, we would not want to say that all such cases involved playing God. Some cases are too insignificant, such as deciding to walk or to drive my car to the shop. Before having a car there was no decision to make. The scale of potential consequences, particularly harmful ones, matter. The Wivenhoe Dam example is probably somewhere in the middle.

6.7 Examples

The field of geoengineering provides some useful examples and the Australian edition of Clive Hamilton's recent book on the topic (*Earthmasters* 2013) is subtitled *Playing God with the Climate*. Geoengineering is being discussed seriously because of growing concerns about climate change and the difficulties of gaining international cooperation in dealing with it. Various strategies are being considered but one of the most attractive, because it would be relatively inexpensive, is solar radiation management, also called albedo modification. The most commonly discussed method is putting sulphate particles or other chemicals into the upper atmosphere to stop some of the sunlight from reaching the Earth. Depending on the amount of particles sprayed into the atmosphere, the increase in temperature could be slowed, or the temperature held stable or lowered. According to Ken Caldeira, a climate scientist at the Carnegie Institution for Science at Stanford,

> We know this is possible because we have seen volcanoes do it. In 1991, the Mount Pinatubo eruption in the Philippines injected lots of small particles high in the atmosphere, and the next year the Earth cooled, despite the continued rise in atmospheric greenhouse gas concentrations. (Caldeira 2015)

So not only would this process be relatively inexpensive but it would also have an effect quickly. While these two features make it attractive it is also highly controversial. Caldeira himself admits:

> It is possible, of course, that sustaining the kind of aerosol layer that circled the Earth in 1991 would just make things worse. We just don't know. (Caldeira, 2015)

However, Caldeira and other advocates such as David Keith (see O'Donnell 2013) view the uncertainty about risks not as showing that the idea should be abandoned but rather that research is required to discover exactly what the risks are and how they can be overcome. The impetus is that climate change could be so disastrous that something drastic may have to be done. (For a detailed discussion of solar radiation management, see Hamilton 2013, ch. 3; MacMartin, Caldeira, and Keith 2014).

Solar radiation management, if it ever takes place, is a good example of humans making decisions in an area where they have little or no expertise and where decisions have not been made before. Cooler or warmer temperatures on a global scale were in the realm of chance, not choice. Some worry that droughts could result from this geoengineering, and these events too have been chance events not caused by human intervention. Those who make the decisions about the amount of sulphate particles to pump into the atmosphere perhaps could be subject to litigation if food shortages result from, or are believed to result from, geoengineering.

In order not to travel well-trodden territory with respect to a human enhancement example, we will briefly consider a fairly radical technique: head transplant (Thomson 2015) or body transplant (Sample 2015) (both describe the same research but from different perspectives). This has been proposed recently, not as an enhancement but rather as a therapy for people who have lost, are losing, or never had substantial use of their bodies. However, if it ever became a reality it would almost certainly be sought as an enhancement. I could be an enhanced human with a much better body (deemed better by me at least) even though my previous body was not lacking relative to other bodies. This research is in a relatively early stage and its feasibility has been challenged, but it is a useful example. If a technique for altering a person can be considered to be playing God in the sense defined here, this would seem to be it. We have never been able to choose our bodies. I am assuming that it makes more sense to say the 'we' attaches to our heads to a greater extent than it does to our bodies. Modifying our bodies is, of course, possible, through prosthetics, various implants, and sex-change operations, for example, but having a whole new body (excluding the head) would be quite new. I can already choose to alter my body; it could become stronger, more tanned, more feminine or masculine, and so on, but only within fairly strict limits. Those limits would be gone with a whole-body transplant. I could choose the race and gender (and perhaps species) of my new body. This would raise some interesting questions, particularly concerning self-identity. If I, as a white Australian, had an indigenous Australian body, how would I see myself and how would others see me? Or as a male if I had a female body? The situation could be uncertain even if I had a different oldish, white, male body from the one that I have now. My head is probably more important than the rest of my body for my identity but that does not mean that my body is unimportant. The situation would be even more interesting if identity could be attached to either head or body. This might be possible if something like the highly contentious idea of cellular memory—that is, 'that human body cells contain clues to our personalities, tastes, and histories, independently of either genetic codes or brain

cells'—turned out to be true (Carroll 2015; see also Cosier 2015). When my body became diseased or too old, I could get a better one, and when my head did, that in turn could be replaced. Perhaps I would be immortal, or at least live as long as a suitable supply of heads and bodies was available. I would be like the axe that was said to be still the same axe after multiple new heads and handles.

This is not to argue that such transplants are morally wrong, although they may be; nor to suggest that they are repugnant, which again they may be. It is simply to point out some of the unknowns.. If it were to become a reality it would definitely expand the area over which humans had control. Our bodies would be matters of choice and not chance to a much greater degree than now. Given the many unknowns, it is not at all clear that we have the competence to make good decisions, and this would be a case of playing God as it is defined here.

Playing God, we said, is expanding the sphere over which we have control to the extent that we no longer have the knowledge or capacity to competently make decisions with respect to the expanded part. 'Competently' is an important qualification because it points to decisions that have significant consequences. Playing God then is a matter of degree for two reasons: competence is not an all-or-nothing issue and consequences can be more or less significant. One can then play God to a greater or lesser extent. Playing God can also change over time. It may be that where we cannot now make competent decisions, with more experience we can. Perhaps one day we will know enough to be able to competently choose our bodies. If that time arrives, body transplants would not constitute playing God.

6.8 Conclusion

So what is wrong with playing God, in the secular sense of interfering with nature? In itself, nothing. The argument has been that regardless of the interpretation of nature, no strong case can made that some human actions are wrong because they interfere with or modify nature. If nature includes people—that is, nature(I)—then everything that we do is as natural as what birds and beavers do. Even if it excludes people—that is, nature(E)—much of what humans do that is necessarily for survival interferes with nature so cannot be morally wrong. This conclusion is at odds with that of van den Belt who, as we saw earlier, says that the wrongness of Frankenstein's action was that he created a *human*. Coady's position is similar to that of van den Belt, concluding that the charge of playing God may be 'an intrinsic moral objection to the very idea of changing human nature' (Coady 2009, p. 179).

It does not follow, of course, that all actions that interfere with nature are sensible, good, or useful, and with respect to the playing God charge, reconsidering Frankenstein is instructive. The wrongness in Frankenstein's action lies not in the creation itself but rather in the fact that he did not think carefully enough about what he was doing. He was acting outside his area of competence. People, at least most, need friends, and Frankenstein's creation had none. If a friend had been created or if the monster had

looked less frightening to humans, there might have been no problems. The monster did, after all, try to be friendly to people at the beginning. His constant rejection led to the problems.

Playing God in the sense of changing or interfering with nature is changing nature in significant ways that change the parameters within which we make decisions such that we are not competent to make those decisions, and where new responsibilities come into play. While there is nothing wrong with this in itself, extreme care must be taken to consider the possible harmful consequences that we may have trouble controlling. This suggests that, strictly speaking, the 'playing God' phrase is redundant and perhaps just a rhetorical device (see Grey 2012). If all of the work is done by the consequences of an action and not by its intrusion into the realm of God or nature, then playing God adds nothing to cost–benefit analysis, the precautionary principle, or some other risk-assessment strategy. From a secular perspective then the notion of playing God is redundant. Any force that it has comes from drawing attention to potential dangers of making decisions and the performance of actions too far outside our areas of competence, but this could be done without talking of playing God. Strictly speaking this is true, but saying that some action is, or might be, playing God draws attention in a dramatic way to the seriousness of potential consequences, and our lack of competence and knowledge. We are familiar with the story of Frankenstein and his monster, and it is worthwhile remembering the *real* Frankenstein problem.

References

Baker, Lynne Rudder (2008). 'The Shrinking Difference between Artifacts and Natural Objects', *APA Newsletter on Philosophy and Computers* 7(2): 2–5, <http://people.umass.edu/lrb/files/bak08shrM.pdf>.

Caldeira, Ken (2015). 'One Known Way to Cool the Earth: Another View', *USA Today* 15 February, <http://www.usatoday.com/story/opinion/2015/02/15/megadrought-climate-change-ken-caldeira-editorials-debates/23465975/>.

Carroll, Robert T. (2015). 'Celluar Memory', in Robert T. Carroll (ed.), *The Skeptic's Dictionary*, <http://skepdic.com/cellular.html>.

Coady, C. A. J. (2009). 'Playing God', in Julian Savulescu and Nick Bostrum (ed.), *Human Enhancement*. Oxford: Oxford University Press, pp. 155–80.

Cosier, Susan (2015). 'Could Memory Traces Exist in Cell Bodies?', *Scientific American*, 9 April, <http://www.scientificamerican.com/article/could-memory-traces-exist-in-cell-bodies/>.

Dworkin, Ronald (2000). 'Playing God: Genes, Clones, and Luck', in Ronald Dworkin (ed.), *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge, MA: Harvard University Press, pp. 427–52.

Flannery, Tim (2003). 'Beautiful Lies: Population and Environment in Australia', *Quarterly Essay* 9: 1–73, <http://caepr.anu.edu.au/sites/default/files/Altman/Reviews/2003_AusAbStudies.pdf>.

Grey, William (2012). 'Playing God', in Ruth Chadwick (ed.), *Encyclopedia of Applied Ethics*, 2nd edn. London: Academic Press, pp. 468–73.

Hamilton, Clive (2013). *Earthmasters: the Dawn of the Age of Climate Engineering*. New Haven, CT: Yale University Press.

Jonas, Hans (1973). 'Technology and Responsibility: Reflections on the New Tasks of Ethics', *Social Research* 40(1): 31–54.

Jones, Rhys (1969). 'Fire-Stick Farming', *Australian Natural History* 16(12): 224–8.

Lewis, C. S. (1960). 'Nature', in C. S. Lewis, *Studies in Words*. Cambridge: Cambridge University Press, pp. 45–6.

MacMartin, Douglas G., Caldeira, Ken, and Keith, David W. (2014). 'Solar Geoengineering to Limit the Rate of Temperature Change', *Philosophical Transactions A* 372(2031), <http://rsta.royalsocietypublishing.org/content/372/2031/20140134>.

Medicinenet (2012). 'Definition of Frankenfood', *MedicineNet.com,* <http://www.medicinenet.com/script/main/art.asp?articlekey=24845>.

Mill, John Stuart (1904). 'On Nature', in *Nature, The Utility of Religion and Theism*. Rationalist Press (from Lancaster e-text, prepared by the Philosophy Department at Lancaster University), pp. 7–33.

O'Donnell, Erin (2013). 'Buffering the Sun', *Harvard Magazine* July-August, <http://harvardmagazine.com/2013/07/buffering-the-sun>.

Ortega Y. Gasset, Jose (1961). 'Man the Technician', in *History as a System: and Other Essays Toward a Philosophy of History*, translated by Helene Weyl. New York: W. W. Norton & Company, pp. 87–161.

Passmore, John (2010). 'Attitudes to Nature', in David R. Keller (ed.), *Environmental Ethics: the Big Questions*. Chichester: Wiley-Blackwell, pp. 103–10.

Peters, Ted (2007). 'Are We Playing God with Nanoenhancement?', in Fritz Allhoff, Patrick Lin, James Moor, and John Weckert (eds), *Nanotechnology: the Ethical and Social Implications of Nanotechnology*. Hoboken, NJ: Wiley, pp. 173–83.

Plumwood, Val (2006). 'The Concept of a Cultural Landscape: Nature, Culture and Agency in the Land', *Ethics & The Environment* 11(2): 1085–6633.

Prince of Wales (1998). 'Genetically Modified Food', Organic Consumers Association, <https://www.organicconsumers.org/old_articles/ge/princewales.php>.

Sample, Ian (2015). 'First Full Body Transplant is Two Years Away, Surgeon Claims', *The Guardian* 25 February, <http://www.theguardian.com/society/2015/feb/25/first-full-body-transplant-two-years-away-surgeon-claim>.

Sandel, Michael (2004). 'The Case against Perfection', *The Atlantic* April, <http://jrichardstevens.com/articles/sandel-genetics.pdf>.

Shelley, Mary (1818). *Frankenstein*.

Thomson, Helen (2015). 'First Human Head Transplant Could Happen in Two Years', *New Scientist*, 25 February, <http://www.newscientist.com/article/mg22530103.700-first-human-head-transplant-could-happen-in-two-years.html?full=true#.VVq_g2Z4ITE>.

TWS (The Wilderness Society Australia Incorporated) (2015). 'What is Wilderness?', *The Wilderness Society*, <https://www.wilderness.org.au/articles/what-wilderness>.

van den Belt, Henk (2009). 'Playing God in Frankenstein's Footsteps: Synthetic Biology and the Meaning of Life', *Nanoethics* 3(3): 257–68.

Wickson, Fern (2008). 'What Is Nature, if it's More than Just a Place without People?', *Nature* 456(29): 6, <http://www.nature.com/nature/journal/v456/n7218/full/456029b.html>.

7

Conservative and Critical Morality in Debate about Reproductive Technologies

John McMillan

7.1 Introduction

Critics of reproductive technologies and human enhancement often object to the hubris of attempting to improve our nature. Unpacking precisely what is meant by such worries is not a straightforward task. I will suggest that one plausible way of reading such objections is that they are in fact concerns about the damage to important 'bearers of value' or 'value institutions' that we risk by departing from what it is natural or ordinary for us to do. I will show this by demonstrating the similarities between the claims of recent moral conservatives about reproductive technologies and those arguing in the middle of the twentieth century. These conservative positions share a common worry about the ability of technology to change or eliminate value-bearers. In the case of assisted insemination, this was expressed as a worry about what this departure from nature might mean for the institution of marriage. That kind of worry is similar to some other recent debates, such as the moves to legalize gay marriage.

Showing how worries about altering value-bearers is at the heart of some conservative objections will enable me to outline a conception of what it is that moral conservatives wish to 'conserve'. Many of those arguing against conservative positions on enhancement, assisted insemination, and gay marriage can be described as believing in a variety of forms of 'critical morality'. This, I will suggest, is the idea that our moral thinking can, and should, be radically revised, even when that means institutions such as marriage might be permanently altered.

Gerald Cohen (2011) developed an account of value conservatism which attempts to explain why it is that we think some particular institutions or 'bearers of value' have value which merits consideration in virtue of being the institutions or objects that they are.

There are different ways in which someone might be conservative about moral rules, principles, value institutions, or bearers of value, but the sense that I'm interested is

Cohen's. He defines value conservatism as 'a bias in favor of retaining what is of value, even in the face of replacing it by something of greater value'. Value conservatives tend to have views that are opposed to claims that we should introduce or embrace change and revise our common views when that will lead to a better state of affairs, a position that I will describe as 'critical morality'.

So, by critical morality I mean the view that moral rules, principles, and value institutions, indeed any bearer of value, can and should be altered in the light of rationality or moral evaluation. It would be a mistake to think that utilitarianism is the only 'critical' moral theory. A virtue theorist might also think that some practices or things that people do fail to conform to what a virtuous person would do in that situation. Likewise Kantian-inspired moral positions might generate duties that contradict what common morality requires. But in saying that, direct utilitarian theories do not value the requirements of common morality if they do not maximize utility, and they are therefore the most starkly 'critical' of moral theories.

I will argue that Cohen's value conservatism adds another value dimension to debates about reproductive technologies and enhancement. While there are problems in knowing exactly what follows from conserving value, it does focus argument on the particular bearers of value that are at stake in a debate and could provide some common ground for critical and conservative moralists.

7.2 Conservativism in Debates about Assisted Insemination and Reproductive Technologies

Debates during the 1930s to 1950s about artificial insemination covered a range of concerns and issues that are quite different from contemporary concerns about assisted reproduction. We take for granted the now settled terms 'artificial insemination' (AI) and artificial insemination by donor (AID), but when these techniques were first transferred from their development in animal husbandry to assist human reproduction it was far from clear what they should be called. One rejected name was 'insemination by a foreign donor', which was viewed as inappropriate because 'the word "foreign" has various connotations and might be used to indicate, say, an Italian, Chinese, or other national'. And for some time the term 'X-insemination' was used so as to highlight the anonymity of the donor (Schellen 1957, p. 4).

While masturbation is still viewed by some religions as sinful, it would be odd if a contemporary debate about a new reproductive technology mentioned that as a reason why it should be impermissible. But, as Glanville Williams observed, that AI would involve the donating male masturbating and collecting his ejaculate was an issue:

> Masturbation is not only a sin, but so sinful that it cannot be permitted whatever good results may follow from it. That masturbation is sinful, when it is a means of obtaining venereal pleasure without the prospect of fertilization, follows, of course, from Augustinian doctrine. (Williams 1958, p. 131)

Intriguingly, some Catholic scholars considered whether there could be a doctrine of double-effect defence:

> Physicians, but not the husband himself, may induce the discharge of the fluid by exercise of the prostate gland, for purposes of either fertility examination or artificial insemination. He allows it *analis frictio* […] It just happens that nearby are the two seminal vesicles, so close that massage of the gland will in fact result in the discharge of fluid. This result, since it is not intended, is legitimate, especially since, he claims, there is no venereal pleasure involved. It is produced only as a secondary consequence of the prostate treatment. (Fletcher 1955, p. 111)

While there were plenty of objections to AID that now strike us as odd, some of the debate was about issues that are familiar to contemporary debates about assisted reproductive technologies. The Church of England commissioned a report on the ethics of AID in 1948. It includes some uneven argument but also a number of more cogent 'conservative' arguments that could well have been developed by contemporary conservatives. Some of them are concerns that you could imagine contemporary conservatives might well have but perhaps not use in debate. For example, they claimed:

> The conscious or ostensible motive may be compassion for the sterile couple, while other motives are present (although unrecognized), such as the desire for power and self-assertion, or the presumption that the donor possesses capacities of such outstanding quality that they ought to be propagated. (Church of England 1948, pp. 27–8)

There's something refreshing about this speculation about the motives of male donors, and perhaps also the ethical duty to consider the purposes to which one's vital fluids are being used. The principal concerns of the report were about what the Church of England thought AID would mean for the institution of marriage. It was less exercised about the sinfulness of masturbation than Catholic commentators, although it clearly thought it immoral:

> The act which produces the seminal fluid, being in this instance directed towards the completion (impossible without it) of the procreative end of the marriage, loses its character of self-abuse. It cannot on this view, be the will of God that a husband and wife should remain childless merely because an act of this kind is required to promote conception. (Church of England 1948, p. 47)

While most would baulk at the description of masturbation as 'self-abuse', this is a more intelligent and creative way to argue about Christian ethics than some of the Catholic discussion. What it is that AID aims at—ameliorating unwanted childlessness—is acknowledged and taken into account in evaluating the morality of what's required for it to succeed. The report extends the action analysis to the question of whether AID implies adultery and whether it is impermissible:

> adultery is not wrong because people enjoy it. The enjoyment is sinful because the adultery is wrong. The action itself, not its accompanying pleasure, determines its moral quality.… Fornication, adultery and (ordinarily) masturbation are wrong because they are all disordered uses of the sex function. (Church of England 1948, p. 50)

At this point in the report the reader might think that therein lies an argument for permitting AID. If the wrongness of adultery and masturbation are only due to their violating the procreative function of reproductive organs, then because AID aims to create a child when there otherwise would not be one, they would appear to be a different way in which sexual organs are properly employed. The report suggests that AI that does not involve donor sperm might be permissible, but using donor sperm is a step too far:

> In the end, however, the judgment we reach on the subject of AID will be shaped and determined by our view of the nature of marriage. If, with the Church, we hold it to be that exclusive contract and union 'for better or worse', the total commitment till death of each spouse to the other, then the part of the unknown donor must be seen as an unlawful intrusion, and the part of the wife—however innocent she may be of evil intent or motive—a breach of her marriage vows… AID extracts procreation entirely from the nexus of human relationships, in or outside of marriage. (Church of England 1948, pp. 56–7)

Perhaps not surprisingly, the woman's infidelity takes centre stage while whether or not the donating male has also committed a sin is not determinative. However, the point that is relevant for this discussion is the potential for AID to change, or perhaps more accurately debase, the value institution of marriage. It's even broader than marriage in that the determinative objection to AID is also about alienating procreation from human relationships. That kind of objection is, at heart, conservatism about how things happen naturally. The Church of England was well aware that some married couples might forgo becoming parents because of this (for better, for worse), but procreation occurring, if not within marriage at least within a context of human social relationships, is not something that should change.

That conservative intuition is very similar to that expressed by more recent critics of reproductive and enhancement technologies. Leon Kass is concerned that

> the road we are travelling leads all the way to the world of designer babies—reached not by dictatorial fiat, but by the march of benevolent humanitarianism, and cheered on by an ambivalent citizenry that also dreads becoming the last of man's man-made things. Make no mistake: the price to be paid for producing optimum or even only genetically sound babies will be the transfer of procreation from the home to the laboratory. Such an arrangement will be profoundly dehumanizing, no matter how genetically good or healthy the resultant children. (Kass 2002, p. 7)

So for Kass, even though there appear to be sound humanitarian and benevolent reasons for developing and making new reproductive technologies available, changing the nature and context of procreation is too important for them to be permissible. Kass does not say whether this is something of such great value that there are no considerations that would make changing the nature and context of procreation justifiable, but in any case his view is an instance of moral conservatism similar to that expressed in the Church of England's report.

Michael Sandel refers to, and endorses, Habermas' worries about designer children. He says that gene therapies that select or improve children are objectionable because of what they entail for autonomy and equality:

> It violates autonomy because genetically programmed persons cannot regard themselves as 'the sole authors of their own life history.' And it undermines equality by destroying 'the essentially symmetrical relations between free and equal human beings' across generations. (Sandel 2007, p. 80)

There are a number of responses that could be made, and have been, to this claim. None of us is the sole author of our life history because none of us had a say over where, how, and to whom they were born, all properties of our life histories. A second problem with the autonomy claim is that it is not as if gene therapies preprogramme individuals—that's a naïve oversimplification of how genes work. But if we focus on Sandel's second claim about equality, it is an instance of moral conservatism. Symmetrical relations between human beings across generations is not obviously a 'value institution' in the sense that I have been using that concept. Rather than a social institution that matters for value, such as marriage, a university, or the law, intergenerational symmetry seems more like a fact about our nature that some think is important for value. It's not hard to think of other natural artefacts that value could be attached to: the distinctive profile of the Matterhorn, a rare bird species, or a river system that has not been dammed. Nonetheless it does seem reasonable to suppose that Sandel and Habermas are concerned about the permanent alteration of something that has existed over time, irrespective of the implications for human welfare or some other morally relevant property. So the class of 'bearers of value' is broader than institutions; it also includes artefacts that have not been created by human beings but which are judged to be of significant or profound value.

7.3 Critical Morality in Debates about AID and Reproductive Technologies

As well as demonstrating familiar conservative arguments, the early ethical debate about AID included views that sought to recalibrate our moral thinking and value institutions for the sake of a higher end. Before the Second World War, a number of respectable biologists and public intellectuals endorsed and argued for eugenics. Prominent thinkers such as Julian Huxley and Nobel laureate Hermann Muller saw the eugenic possibilities of AID and said we should rethink our views about marriage and procreation:

> artificial insemination has opened up new horizons, by making it possible to provide different objects for two functions. It is now open to man and woman to consummate the sexual function with those they love, but to fulfil the reproductive function with those whom on perhaps quite other grounds they admire. This consequence is the opportunity of eugenics. But the opportunity

cannot yet be grasped. It is first necessary to overcome the bitter opposition to it on dogmatic theological and moral grounds. (Huxley 1943, quoted in Church of England 1948, p. 1)

For Huxley, insisting that procreation must occur within the context of marriage forgoes the possibility of benefiting humankind. This example of critical morality turns on the importance of revising a value institution for the sake of a morally important end. Huxley's contemporary and collaborator, Hermann Muller, wrote a utopian novel which describes a future society where procreation and marriage have become disentangled:

'eugenic marriages' cannot as a whole be successful, so far as the parents are concerned. On the other hand, make personal love master over reproduction, under conditions of civilisation, and you degrade the germ plasm of the future generations. Compromise between the two policies, and you cripple both spirit and sperm. There is only one solution—unyoke the two, sunder the fetters that from time immemorial have made them so nearly inseparable, and let each go its own best way, fulfilling its already distinct function. (Muller 1936, p. 139)

Muller was convinced that by creating sperm banks stocked with the gametes of 'superior' men and having a programme of mass donor insemination, humanity could be improved within a few decades (Paul 1987, p. 324). Initially his claims were primarily about the importance of negative eugenics, but by the mid-1930s he was pushing hard for positive eugenics, even to the extent that he sent a copy of *Out of the Night* to Stalin and attempted to convince him that this was a natural development for socialism (Paul 1987, p. 326).

While the technologies being debated are different, there is a similarity between this attempt to increase the frequency of positive non-disease states and the principle of procreative beneficence defended by Julian Savulescu:

we have a moral obligation to test for genetic contribution to non-diseases states such as intelligence and to use this information in reproductive decision-making because intelligence contributes to the quality of person's life no matter what their plan of life. (Savulescu 2001a, p. 415)

It might be objected that the principle of procreative beneficence is a liberty respecting principle: agents should be free to make the procreative choices they wish, subject to constraints such as the harm principle (Robertson 1994). Moreover, this is an obligation that aims to produce the best children that parents can and not to improve the genetic health of future generations, so it is not promoting positive eugenics. However, the differences are not as sharp as they might appear. Huxley and Muller also believed that this is a moral requirement that those procreating would recognize once they were informed about the benefits of selective breeding. Likewise they thought that people should be free to make their own reproductive choices, so this is a matter of morality rather than permissibility. While Savulescu seems principally concerned with the children who will immediately result from a procreative choice, he is in fact endorsing a positive eugenic view because the benefits he thinks will flow to the immediate offspring should also be benefits to subsequent generations.

While Savulescu mentions intelligence as a trait that will benefit and enhance human beings, irrespective of their conception of the good (2001b), Muller adds a trait that has been debated more recently as a possible moral enhancement for human beings:

> The whole course of human prehistory and history makes the necessary ideals obvious and unmistakable; they are (aside from physical well-being) primarily two: highly developed social feeling—call it fraternal love, or sympathy, or comradeliness, as you prefer; and the highest possible intelligence—call it analytical ability, or depth of understanding, or 'reason'. (Muller 1936, p. 148)

Muller means that we should positively select character traits that will reduce violence and aggression while increasing cooperative behaviour. Although he doesn't use Persson and Savulescu's preferred term 'empathy' (2011), it's clear he thinks he's referring to something quite similar and that selecting for this attribute would benefit humankind.

For the purposes of this discussion there's no need to pigeonhole the views of Savulescu and others who argue for the principle of beneficence as utilitarian or any other moral theory. What is significant is the way in which he wants us adopt a new set of moral obligations and revise our standard way of viewing the morality of our procreative choices. This approach to practical ethics is part of a fine tradition that includes thinkers such as Jeremy Bentham and Peter Singer, who can be interpreted as directing their intellectual effort towards revising legislation and common morality so that it tracks the requirements of morality more directly. That's a tradition that eugenicists such as Muller and Huxley would have been happy to join. Herbert Brewer coined the term 'eutelegenesis' (1935) to refer to mass AID for the purposes of positive eugenics, and he quite clearly sees himself as bringing rational requirements to bear on conventional morality:

> Conventionality is not the same thing as morality. Morality goes far beyond following already approved folk ways. The highest morality consists not in uncritical adherence to old customs and beliefs, but in seeking to discover what new ones harmonize with objective truth and serve best the changing needs of progressive civilization. (Brewer 1937, p. 344)

Like critical moral philosophers who think we should reshape morality to fit what some independent account of the right implies, Brewer and other early eugenicists thought that there could be an objective truth about what would best suit the needs of humanity and we should revise common morality towards that end.

We might want to disagree with what the early eugenicists claimed is possible via selective breeding. Some will also worry about treating the moral rules, principles, and institutions that we act on as all being up in the air and readily reconstructed in the light of a critical morality. However, those such as Kass, Sandel, and the authors of the Church of England's report would appear to have a much worse problem. They claim that procreation within a loving relationship, and symmetrical relations between

generations and marriage are valuable, yet none of them give us much in the way of an argument for why they are of value and why we have reason to conserve these things when changing them might lead to some other morally worthy end.

One of the better and more interesting defences of such conservative intuitions is that developed by Gerald Cohen (2011). In Section 7.4 I'll explain how Cohen's account works and consider whether it might be used to decide when and how much weight we should give to conservative intuitions when arguing about reproductive technologies.

7.4 Defending the Conservative Intuition

Cohen's central aim is to explain the conservative bias in favour of retaining bearers of value even when they might be replaced by something of greater value. If that account works then it might be used to help adjudicate and weigh appeals to existing value, such as those made by Sandel and Kass. Cohen begins his account with a discussion of Kenora College, a fictional North American undergraduate college, which is considering whether it should offer a doctoral programme. Those in favour of changing the college present arguments about the benefits to the institution, including that a doctoral programme is likely to enhance the college's undergraduate programme. Cohen describes the objection that a 'conservative' professor makes to the proposed changes:

> I do not *dis*believe that undergraduate education would be promoted by the proposed changes. Rather, I challenge a presupposition of the young professors' argument, which is that everything that we justifiably decide to do may and must be justified as conducive to some good that our decision might produce. For, in addition to the consideration of what good we might do, which must of course affect our decisions, there is also the consideration of what we *are*, of our identity, and we may legitimately have regard to our desire to preserve that identity. (Cohen 2011, p. 5)

The conservative professor's appeal is to the status of the college being an undergraduate college because it is a fundamental part of the college being the college it is. So, while there might be advantages to changing the nature of the college, conserving it as the institution it is, is a significant countervailing consideration.

He claims that the college is in this case a bearer of intrinsic value, and this value is something that all agents have to reason to see. That seems plausible: those with no connection to the college are likely to feel less strongly about its preservation, but others should be able to see the force of the conservative intuition. This means that there are a couple of reasons why Cohen's defence of conservatism could be of use for debates about reproductive technologies and enhancement. His conservatism is not subjectivist: whether or not a value-bearer is worthy of conservation is not simply a matter of what one person believes and is only true for that person. The conservative professor is not merely stating a preference of feeling that she has towards the college; her conservative argument is intended to have rhetorical force and to be convincing to those with a different view. That's also a feature of the conservative arguments of Kass, Sandel, and

the Church of England's report. They're not merely describing why marriage and procreation are valuable to them in their present form but also suggesting that others should understand their value. This is a positive thing for those wishing to engage with, or critique, conservative positions in that this form of conservatism is amenable to reason.

A second reason why Cohen's conservatism is helpful for keenly contested issues is that in addition to not being subjective, the professor's argument about Kenora College is truth apt. She appeals to a number of factual considerations that could be true or false. If an opponent discovered that in fact the rosy view the college had of its undergraduate education was not warranted, perhaps because the professors were not up to speed due to having no graduate students, then the value of conserving the college as it is would be undermined.

By drawing attention to the significance of some bearers of value being that particular bearer of value, Cohen shows a way forward with debate about apparently intractable issues. For example, if we think that many of the ends in ethical debate are 'ultimate ends'—that is, ones for which no further justification can be given—then the modes of argument at our disposal are limited. If someone objects to the felling of a large, old but not ancient or rare tree that a city council wishes to remove because it places children at risk, there are number of argumentative moves we could make. We might invite that person to think about cases where we might remove hazards to children and ask them whether they really think the tree is more important than eliminating the risk. But if they offer no reason for the value they attach to the tree and simply say that it just is valuable and should not be cut down, then we are at an impasse with respect to the argument. Of course, the council might have the power to just fell it and do so anyway, but a decision that responds to the nature of the objection and takes it into account cannot occur if no further justification for the conservative intuition is given.

If conservatives have to account for why a particular bearer of value is of value, then that increases the cogency of their argument, it creates the possibility that a reasoned solution could be met, and it provides opponents with a way of defeating conservative intuitions. In the tree case, someone could defend such an appeal by referring to facts such as the age of the tree, who planted it and why, its historical significance, and so on. Facts of that kind could justify and provide grounds for argument about its value.

However, there are at least two significant problems with arguments that appeal to the conservative intuition. The first is about the extent to which a value-bearer can change before its nature is changed in a significant way, and the second about knowing how much value to grant a value-bearer.

The Church of England's report ended up rejecting AID because it meant that marriage vows would be broken and marriage is too important, as it is, to be revised in this way. The nature of marriage vows could, presumably, be reinterpreted so that AID did not imply a breach, but that flexibility was not a change that the Church of England was willing to accept. The report doesn't say much about why marriage, as it is, is a bearer of value, but it's likely that its history and development within the Church, its links to

scripture, and its role in constituting a specific conception of the family and related assumptions what that means for the welfare of children, and so on are part of the story. An elaboration of that kind would have made the Church of England's rejection more cogent in that those who are not Anglican would have had reason to take it seriously, and it would also have meant we had a way forward in thinking through whether AID would change the nature of marriage.

The report might be correct that being more flexible about marriage so as to accommodate AID would imply changing marriage, but is that a change that alters the nature of marriage? Conservatives cannot and should not object to any change to a value-bearer because they inevitably do change over time. Marriage has developed and been refined over the course of history, and that is part of what might be thought to make it a bearer of value. In the case of Kenora College, the conservative professor claims it is essential to the college being the college it is that it remains without graduate students. An argument could be mounted for this being a property the college has always had or additionally that it was specifically founded for that purpose, and we might be convinced by that. On the other hand, an obvious response to her is to say that the college will change, but change is inevitable and because it will be in the same place, have the same name, and so on, there is no reason to think that the college, as it is, would no longer be there.

So while the conservative intuition appears to help with argument, there is still an area of vagueness around what is essential to a bearer of value.

Cohen doesn't (perhaps because no one could) spell out how much value conserving something should be accorded. We might all agree with the Church of England's report and the Kenora professor that enabling more to procreate and increasing the wealth and scholarship of Kenora College are not the only values in play, but that raises the difficult issue of knowing how important it is to conserve such bearers of value. It might be that we think changing marriage laws so that gay and lesbian couples can marry does indeed change the nature of marriage and that such changes count for something, but not very much, certainly not enough to prevent lesbian and gay couples from having that right. Kass' worry about our changing the nature of marriage or procreation seems subject to a similar kind of problem: we might all agree that new reproductive technologies change the nature of something important, that this counts for something, but think it's really not important enough to outweigh the reasons why it's important. Kass, on the other hand, is likely to attach a great value to it, even if it means that important life plans go unfulfilled.

7.5 Conclusions

The two problems I raised for arguing about the conservative intuition are drawbacks, but they're not unique to conservatism. Making a judgement about whether the harm that results from an agent exercising a liberty is sufficient for curtailing their liberty is a common moral and legal balancing act for which we tend to rely on history and analogy

for guidance. When it comes to how much weight conserving should be given, we might argue in a similar way by comparison with other cases.

Small enhancements to our ways of arguing about practical ethics should be welcomed, and Cohen's justification of the conservative intuition might in some cases take us beyond a stand-off between conservative and critical moralists.

References

Brewer, Herbert (1935). 'Eutelegenesis', *Eugenics Review* 27: 121–6.

Brewer, Herbert (1937). 'Eutelegenesis and Sterility', *Eugenics Review* 28: 344.

Church of England, Archbishop of Canterbury's Commission on Artificial Human Insemination (1948). *Artificial Human Insemination: the Report of a Commission Appointed by His Grace the Archbishop of Canterbury.* London: SPCK.

Cohen, Gerald (2011). 'Rescuing Conservatism', in R. Jay Wallace, Rahul Kumar, and Samuel Freeman (eds), *Reasons and Recognition: Essays on the Philosophy of T M Scanlon*. Oxford: Oxford University Press, pp. 1–39.

Fletcher, Joseph (1955). *Morals and Medicine*. London: Victor Gollancz.

Huxley, Julian (1943). *The Uniqueness of Man*. London: Chatto & Windus.

Kass, Leon (2002). *Life, Liberty and the Defense of Dignity: the Challenge for Bioethics*. San Francisco, CA: Encounter Books.

Muller, Hermann (1936). *Out of the Night: a Biologist's View of the Future.* London: Victor Gollancz.

Paul, Diane (1987). ' "Our Load of Mutations" Revisited', *Journal of the History of Biology* 20(3): 321–35.

Persson, Ingmar and Savulescu, Julian (2011). 'Getting Moral Enhancement Right: the Desirability of Moral Bioenhancement', *Bioethics* 27(3): 124–31.

Robertson, John (1994). *Children of Choice.* Princeton, NJ: Princeton University Press.

Sandel, Michael (2007). *The Case Against Perfection: Ethics in the Age of Genetic Engineering*. Cambridge, MA: Harvard University Press.

Savulescu, Julian (2001a). 'Procreative Beneficence: Why We Should Select the Best Children', *Bioethics* 15: 413–26.

Savulescu, Julian (2001b). 'In Defense of Selection for Nondisease Genes', *The American Journal of Bioethics* 1: 16–19.

Schellen, A. (1957). *Artificial Insemination in the Human*. Amsterdam: Elsevier.

Williams, Glanville (1958). *The Sanctity of Life and the Criminal Law.* London: Faber and Faber.

8

Human Enhancement: Conceptual Clarity and Moral Significance

Chris Gyngell and Michael J. Selgelid

8.1 Introduction

The lack of consensus regarding how best to understand 'enhancement' has long been cited as a factor impeding ethical debate about enhancement technologies. In 1998, Erik Parens reported that many bioethicists believed 'the term enhancement is so freighted with erroneous assumptions and so ripe for abuse that we ought not even to use it' (Parens 1998, p. s1).

More recently, Earp and co-authors stated:

> 'enhancement debates' in biomedical ethics have been needlessly encumbered by the existence of a hodge-podge of ill-defined, poorly articulated notions of enhancement—often only implicitly communicated—along with endless to-ing and fro-ing about the relationship between enhancement and the limits of medicine. (Earp et al. 2014, p. 5)

In this chapter we consider different ways in which the term 'enhancement' is used in the bioethics literature. In Section 8.2 we outline seven different definitions of enhancement. We first examine approaches that define enhancement as the beneficial alteration of capacities outside the context of treatment. We then turn attention to accounts that define it as bringing about traits or abilities beyond what is typical or possible for our species. We then discuss more recent attempts to define enhancement that incorporate normative notions such as 'a good life'. Finally we examine the idea that enhancements are interventions that increase functioning.

In Section 8.3 we discuss how the definition of enhancement can help frame the ethical debate about capacity-altering biotechnologies. We argue that the term 'enhancement' is useful in moral and policy debates. Some accounts of enhancement draw attention to particular morally relevant spectra and are useful in the ethical assessment of particular technologies. Other approaches to enhancement help reorient debates about biotechnologies and draw attention to salient, but often neglected, questions.

8.2 Accounts of Enhancement

8.2.1 The Constructivist Approach

Enhancement is perhaps most commonly conceived as the non-therapeutic beneficial alteration of capacities—that is, the beneficial alteration (of capacities) that does not involve treating or preventing disease. Drawing on Eric Juengst (1998), Earp and co-authors describe (but do not endorse) the 'not-medicine' approach, which defines enhancement as 'interventions designed to improve human form or functioning beyond what is necessary to sustain or restore good health' (Juengst 1998, cited in Earp et al. 2014, p. 2).

This approach draws a sharp distinction between enhancements and therapies. The use of respiratory medications to improve lung functioning is a therapy when used by someone suffering from lung disease, but it is an enhancement when used by a healthy individual. This view of enhancement is also reflected in the title of some of the most influential works on human enhancement, including the President's Council on Bioethics report *Beyond Therapy* (President's Council on Bioethics 2003) and Elliot's *Better Than Well* (Elliott 2003).

This approach to enhancement assumes that we have an uncontentious and clear account of disease and health (Resnik 2000). However, what constitutes a disorder or disease is highly contested, and different scholars work with different conceptions of disease. Depending on how one characterizes disease, one can describe different variants of the not-medicine approach to enhancement (Savulescu, Sandberg, and Kahane 2011).

Constructivists about disease believe that classification of conditions as diseases, reflects societies' norms and values (rather than being an objective matter). Different societies with different norms and values will consider different states to be diseases. In support of this claim, constructivists point to cases such as drapetomania, a 'disease' described in 1851 which caused slaves to run away (Cartwright 1851), and the fact that masturbation used to be considered a disease (Engelhardt 1974).

This definition of disease can be used to develop what we will call the Constructivist Approach (CA) to enhancement. Under CA enhancements are individually beneficial alterations to human functions or capacities which are not performed in the context of treating or preventing disease, with diseases being understood as states which are disvalued by society in a particular way.[1, 2]

[1] One criticism of CA is that it is difficult to provide a convincing account of the 'particular' way in which disease states are disvalued, without appealing to objective facts about the body. It is clear that not all states which are disvalued by society are considered as diseases. Being 'impoverished' or 'unpopular' are examples of states which would not be considered as diseases despite being disvalued. Hence proponents of CA need to provide an account of the particular way in which diseased states are disvalued which does not appeal to facts about the body.

[2] This view of enhancement is similar to the 'ideological approach' described by Savulescu, Sandberg, and Kahane (2011).

In CA there are no underlying biological differences between enhancements and therapies. What we happen to classify as enhancements or therapies is determined by the current values of society, which can change over time and/or be different in different societies. Something like this definition of enhancement is reflected in the work of James Canton:

> The future may hold different definitions of human enhancement affecting: culture, intelligence, memory, physical performance, even longevity. Different cultures will define human performance given their social and political values. It is for our nation to define these values and chart the future of human performance. (Canton 2003, p. 78)

CA offers a relativistic account of the disease-enhancement distinction. Which technologies are considered as enhancements, according to CA, is determined by what values society happens to hold rather than being a scientifically objective matter. According to CA, for example, the fact that we believe giving Ritalin to someone suffering from attention deficit hyperactivity disorder (ADHD) is a therapy while giving it to a non-sufferer is an enhancement merely reflects that (our) society especially disvalues the cognitive style associated with ADHD.

8.2.2 The Normal Functioning Approach

Objectivists differ from constructivists in that they believe disease involves some type of objective underlying biological disorder. The most well-articulated approach to the not-medicine approach to enhancement has been derived from the Biostatical Theory (BST) of disease, developed by Christopher Boorse (1977). The BST holds that a disease is something which 'reduces one or more functional abilities below typical efficiency'. The normal function of a given trait within a species is described as its 'statistically typical contribution(s) […] to […] individual survival and reproduction' (Boorse 1977, p. 555). Hence, according to the BST, we all have 'species-typical' biological functions which support our survival and reproduction, and diseases are states in which these functions are compromised.[3] We can determine whether an individual is suffering from a disease statistically by comparing their level(s) of functioning with those of others in an appropriate reference class (normally others of the same species, sex, and age). For example, if someone's cardiac functioning (e.g. as reflected by cardiac output) is abnormally low, we can conclude that they suffer from some form of cardiovascular disease.

Swartz, building on earlier work by Norman Daniels (1994), uses the account of disease provided by the BST to distinguish between therapies and enhancement:

> Philosophers and ethicists have defined 'health' in myriad ways, but the approach that I would favor in this context ties health to normal functioning… The function of a given trait in a

[3] It is the disruption of functions that support survival and reproduction, rather than the disruption of survival and reproduction itself, which constitutes disease according to BST. For example, catching a cold often has no effect on one's survival and reproduction, but it disrupts functions which support survival and reproduction, such as pulmonary function.

species is the way that it typically contributes to survival and reproduction in individuals with that trait . . . Focusing on proper functioning in this way correctly classifies which interventions are treatments and which are enhancements. For instance, if the presence of some toxin (such as lead) in a child's home would interfere with the normal development of his brain, then removing the toxin or counteracting its effects would certainly count as a medical intervention. And this remains true even if the result is raising the child's IQ from 140 to 160. In contrast, if there is no toxin, and the intervention instead raises the IQ from 140 to 160 by causing overgrowth of some part of the brain, this would count as enhancement. (Schwartz 2005, p. 18)

Drawing on this work we can describe a different version of the not-medicine approach to enhancement which uses a normal-functioning account of disease. The Normal Functioning Approach (NFA) sees enhancements as individually beneficial alterations to human capacities which are not performed in the context of treating disease, with diseases being understood as negative deviations from normal functioning.[4]

In sum there are at least two different versions of the not-medicine approach to enhancement—CA, which sees enhancement as essentially subjective and relativistic, and NFA, which sees enhancement as scientifically objective.

8.2.3 Beyond-Species-Typical Approach

Although Daniels' normal functioning account of enhancement was originally formulated to help distinguish therapies from enhancements, it can also form the basis of a different definition of enhancement that is independent of therapy. We will call this approach the Beyond-Species-Typical Approach (BSTA) to enhancement, which was formulated, though not endorsed, by Savulescu, Sandberg, and Kahane (2011, p. 5). BSTA defines enhancements as biological or psychological alterations that increase a person's functioning beyond species typical levels.[5]

Similar to NFA, BSTA refers to what is 'normal' for the species. If an intervention causes an individual's capacities to rise above levels that are statistically typical of the species, it is considered to be an enhancement. To see the difference between BSTA and Daniels' original NFA, consider the following case:

Johnny is a short eleven-year-old boy with documented growth-hormone deficiency resulting from a brain tumor. His parents are of average height. His predicted adult height without growth hormone (GH) treatment is approximately 160 cm (5 feet 3 inches).

Billy is a short eleven-year-old boy with normal GH secretion according to current testing methods. However, his parents are extremely short and he has a predicted adult height of 160 cm (5 feet 3 inches). (Allen and Fost 1990, p. 18)

[4] With normal functioning, in turn, understood as per the BST as a statistical concept.

[5] When species-typical functioning is referred to in the BST, statistical abnormality (constituting disease) is commonly defined as being below two standard deviations below the mean (see e.g. Erk 2011, p. 46). Similarly we will assume that two standard deviations above the mean is the cut-off point for statistical abnormality (constituting enhancement) for the purpose of the BSTA.

Let's imagine that if both Johnny and Billy are given synthetic growth hormone, each would reach a height of 180 cm. According to NFA, giving growth hormone to Johnny is a therapy because his short stature is caused by a disease—a type of malfunction. Conversely, giving it to Billy is enhancement because his short stature is caused by normal processes. However, according to BSTA, neither counts as an enhancement because neither intervention causes Billy or Johnny to rise above species-typical levels. However, if either intervention caused Billy and Johnny to grow much taller, say to 220 cm, both would be considered to be enhancements according to BSTA.

Using species-wide values to calculate what is 'typical' can have counterintuitive implications. Consider the case of female athletes taking testosterone. On BSTA these athletes would generally not be enhancing themselves because their testosterone levels would usually stay below species-typical levels. Conversely, young male athletes who take testosterone will generally be enhancing themselves, according to BSTA, because the intervention usually takes them beyond species-typical levels. It seems odd that these two acts should be classed differently. The fact that there is broad variation in traits between same-species members means that the BSTA will often have counterintuitive implications. As Kingma (2007, p. 128) states, 'the human species shows a wide variety of functioning; what is normal in one group can be abnormal in another'. To avoid this problem, alternate versions of BSTA could define enhancements as interventions that increase a person's functioning beyond those which are typical for other's of the same sex and/or age.

However, some have argued that it is inconsistent to use only sex and age to build reference classes.[6] Ethnicity, place of residence, employment status, and so on all affect what values are typical for any given trait. Why not create reference classes based on these characteristics? What about all other factors that affect our traits? When we start to consider factors beyond species membership, we regress to a highly individuated notion of enhancement. However, if we do this we are creating a definition of enhancement very different from that originally outlined by BSTA.

8.2.4 Beyond-Species-Maximum Approach

A different approach to enhancement but one that still draws on the notion of species is the Beyond-Species-Maximum Approach (BSMA). According to this, enhancements are 'alternations to capacities which take people beyond what is naturally possible for the species'.[7,8]

This is a narrow definition of enhancement. For individuals who are not already at the maximum level for a given trait, only fairly extreme interventions will be considered as enhancements on BSMA. For example, few people who take steroids achieve

[6] The challenge of specifying the members of reference classes applies to the BST as a whole and so is applicable to both BSTA and NFA.

[7] This is similar to how Nick Agar (2010) defines radical enhancement.

[8] One could drop the 'naturally' requirement from this definition to avoid some of the difficulties we discuss below. However, then it becomes doubtful that any interventions will be considered enhancements because none would give people capacities beyond what is possible for the species to achieve.

values in traits that are beyond what it is possible to achieve naturally for *Homo sapiens*. Thus taking steroids is unlikely to be an enhancement on BSMA in most cases. On the other hand, extensive genetic engineering that causes radical changes to our phenotype could be considered as an enhancement because it may cause individuals to acquire values or properties that are not 'naturally' possible for humans.

One problem with using this approach is that it is difficult to determine what the species-maximum values are for particular traits. For a species that currently consists of 8 billion individuals, how can we know what is even currently possible? For example, before Usain Bolt ran 100 metres in 9.58 seconds in 2009, many would have thought that this was not possible. Before 2009, any athlete who used steroids so that they could run as fast as Usain Bolt would have been thought to be enhancing themselves according to BSMA. However, we now know that such feats are possible to achieve 'naturally' (or without the use of drugs anyway),[9] so presumably they would not be considered as enhancements according to BSMA. The fact that we generally don't know what values for traits are possible for the species means that the BSMA is hard to apply.

Relatedly, the notion of species-maximum levels is incoherent under some species concepts. According to one influential view, a species is 'a lineage of ancestral descendant populations which maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate' (Wiley 1978, p. 17). On this view, species are seen as lineages composed of individuals who stand in certain ancestral relations to each other. If we think about species in this way, the notion of species-maximum values seems problematic. *Homo sapiens* is defined as a lineage rather than an organism with particular characteristics. Our traits could radically change through natural evolutionary processes without changing which species we belong to. Hence any properties we could acquire through the use of enhancement technologies would not take us beyond species-maximum values.

A final objection to BSMA relates to its appeal to a supposed natural/unnatural distinction, because it is not always (if ever) clear what this distinction is meant to amount to. If someone consumed an unusually large amount of some nutrient, would that be unnatural? Would it be natural if the nutrient in question was contained in, say nuts, that the person ate (lots of), but unnatural if the same nutrient was produced chemically? Would it matter if the chemical production of the nutrient was achieved with naturally existing chemicals, or would its chemical production need to involve chemicals contrived by humans for the nutrient intake to count as 'unnatural'? Among numerous other worries about the supposed natural/unnatural distinction are that everything we do will be occurring in nature. Human beings and human actions are themselves part of nature, so the fact that human beings do (or make) something surely cannot be sufficient reason to call that thing 'unnatural'.

[9] This chapter was written in late 2015, a time when it was generally accepted that Usain Bolt's performance in 2009 was not aided by the use of steroids.

8.2.5 The Welfarist Conception

Recently there has been a push to understand enhancement in a normative sense. Julian Savulescu and co-authors have developed a detailed normative account of enhancement called the Welfarist Conception (WC) of enhancement. WC defines enhancement as any change in the biology or psychology of a person which increases their chances of leading a good life in a given set of circumstances (Savulescu, Sandberg, and Kahane 2011, p. 7).

Savulescu and co-authors explain:

> On this view, any increase in IQ could count as enhancement – so long as it tends to increase a person's well-being. But, contrary both to the species-functioning and functional approaches, in contexts where an increase in IQ is not beneficial to some person, such increase would not count as an enhancement, even if it raises the person to (or well beyond) the level of normal functioning, that is, even if it were a functional enhancement. (Savulescu, Sandberg, and Kahane 2011, p. 7)

This is a very broad and revisionist notion of enhancement. Many interventions which are not normally considered as enhancements would be classified as such under WC. For example, as treating disease improves well-being, nearly all therapies are considered to be enhancements under WC. Similarly, other things that are not normally considered as enhancements improve people's lives through changes to biology or psychology. Contraceptive pills, for example, generally improve the lives of women by increasing their reproductive liberty. In the WC, then, taking contraceptive pills is an enhancement.

Even basic improvements in diet—for example, reductions in sugar or salt—would end up counting as enhancements, on this view, so long as the dietary changes played a causal role in improving a person's well-being. Though we don't necessarily think there is one objectively correct definition of 'enhancement', we seriously doubt that this is the concept that most users of the term (in debates about the ethics of human enhancement) have in mind.[10]

8.2.6 The Modified Welfarist Approach

WC was originally presented as an alternative to the not-medicine approach to enhancement. In its original formulation WC does not incorporate a distinction between therapies and enhancements. This is because it focuses on 'changes' to biology and psychology; all (beneficial) changes—therapies or not—are treated equally. However, a modified version of the welfarist approach to enhancement can be developed which incorporates an enhancement/therapy distinction.

One way to understand disease is as a *thick* concept, containing both normative and descriptive elements. On this view, when we characterize something as a disease, we

[10] This, of course, could be studied empirically.

are making the value-based judgement that it reduces well-being, and in addition saying that it is caused by some type of biological malfunction (Murphy 2015). Hence if someone has abnormal biological functioning which reduces their chances of living a good life, they are considered to be in a diseased state.[11]

However, not all abnormal biological functioning will be detrimental to one's well-being. Many cases will be neutral, and some will be beneficial. If someone has abnormal biological functioning that increases their well-being, we can describe them as being in an advantaged or enhanced state. Hence we can distinguish the following states based on biological functioning within a WC:

1. diseased state: abnormal biological functioning, which is detrimental to well-being
2. normal functioning: no abnormal biological functioning, or abnormal biological functioning that does not affect well-being
3. enhanced state: abnormal biological functioning, which increases well-being.

This allows us to draw a distinction between therapies, which move people from state 1 to 2, and enhancements, which move people from state 1 or 2 to 3. Hence in this Modified Welfarist Approach (MWA), enhancements are alterations that give people abnormal biological functioning that improves well-being.[12]

Like WC, MWA implies that every enhancement is good for you. However, enhancements must also give individuals some type of abnormal biological functioning.

8.2.7 The Functional Approach

Another way the term 'enhance' can be used is to signify interventions which increase performance on some measure. We might say that caffeine enhances reaction speed, by which we just mean that caffeine increases reaction speed. Similarly, some people use 'enhance' to refer to increases in some capacity. By saying 'taking steroids enhances muscle growth', we just mean that steroids increase muscle growth. This sense of enhancement fits with the definition of 'enhance' in the *Shorter Oxford English Dictionary:* 'To raise in degree, heighten, intensify' (Little, Fowler, and Coulson 1959, p. 612).

In this spirit we can describe the Functional Approach (FA) to human enhancement, which defines enhancements as interventions which increase some function or capacity.[13] In their review piece about enhancement definitions, Earp and co-authors describe a similar approach to enhancement which they call the 'functional-augmentative approach':

[11] We can then define abnormal biological functioning statistically—as per BSTA, or in some other specified way. Disease, in this case, ends up being very similar (or perhaps the same) as what Boorse refers to as illness.

[12] One can also formulate a different modified welfarist conception, which drops the requirement for the alterations to result in abnormal biological functional. This would be slightly different from WC, by focusing on changes to functioning rather than just changes to biology or psychology.

[13] This latter point recognizes that some enhancements may give people entirely new abilities rather than just increasing their existing functions.

> Interventions are considered enhancements insofar as they improve some capacity or function (such as cognition, vision, hearing, alertness) by increasing the ability of the function to do what it normally does. (Earp et al. 2014, p. 2)

In this approach, 'enhancement' is a purely descriptive term, referring only to the direction of change achieved in a particular capacity or functioning. Enhancements always increase functioning. Strength enhancements always make people stronger; memory enhancements always increase memory; and so on. In addition to applying to specific traits such as memory and strength, FA can be applied more broadly to groups of traits. We can say that something is a cognitive enhancement when it increases the functioning of at least one cognitive capacity, and something is a physical enhancement when it increases at least one physical capacity.

One advantage of the FA is that its extension is often clear. The claim that something is an enhancement just means that it increases some capacity, which can often be easily checked.[14] Hence the class of technologies which are 'memory enhancements' or 'strength enhancements' is relatively easy to determine.

One difficulty with FA is that it does not closely align with how the term 'enhancement' is ordinarily used. Many therapies work by increasing some capacity or functioning. For example, respiratory medications increase respiratory functioning, sleep apnoea treatments increase alertness, and cochlear implants increase auditory functioning. On FA, these interventions will be considered as enhancements. However, these are normally considered as therapeutic interventions rather than as enhancements. FA does not respect the commonly made distinction between therapy and enhancement.

Further, some interventions generally considered as enhancements are not captured by FA. For example, consider the use of propranolol to reduce the emotional intensity of certain experiences, or the possibility of 'antilove' biotechnologies. Earp et al. believe that:

> interventions of this kind raise many of the same patterns of ethical concern as the more conventional cases of functional enhancement typically encountered in the bioethics literature… At the same time, however—given a functional-augmentative framework—these cases might seem puzzling or out of place. They seem puzzling because they apparently involve the very opposite of enhancement, namely, diminishment: i.e., diminishment of wartime memories; diminishment of harmful love; diminishment of ill-directed lust, and so on. (Earp et al. 2014, p. 3)

In the view of Earp and co-authors, some accepted forms of enhancement seem to decrease rather than increase capacities or functioning.[15]

[14] Of course, something can be an enhancement in one sense but a diminishment in another sense. Alcohol, for example, may enhance confidence but diminish balance.

[15] Most of the examples used by Earp et al., seem to assume an etiological account of function rather than a goal contribution account, which is favoured by Boorse. Take the example of the diminishment of painful memories. On a goal contribution account of function, increasing the function of memory would entail increasing its contribution to survival and reproduction. Hence the diminishment of painful memories could increase the functioning of the memory system in so much as it helps people live their lives more effectively. Hence the diminishment of painful memories will often be an enhancement in FA, if

In addition, some biotechnologies may not increase or diminish particular capacities but rather give individuals entirely new ones. Imagine advanced biotechnologies which allowed individuals to navigate through the use of sonar, like some species of bats. These technologies would not be considered as enhancements under FA as they do not act on any existing function. However, many people would regard interventions like these as a type of enhancement.

Hence FA is in some ways a revisionist definition of enhancement—it is different from the lay understanding of enhancement, which draws a distinction between treatment and enhancement. FA sees enhancements as just one type of human modification technology. In addition to function-increasing enhancements we must also recognize diminishments—interventions which decrease some function or capacity—as well as biotechnologies that give people entirely new capacities.

8.2.8 Summary of Enhancement Definitions

We have provided seven different definitions of enhancement.[16] These are summarized in Table 8.1, which highlights points of convergence and divergence by noting whether each definition implies that giving growth hormone to Johnny or Billy is an enhancement in the case from Allen and Fost (1990, p. 18) mentioned earlier.

These are all potentially useful definitions of 'enhancement', and we do not believe there is a single correct way to define the term. For the remainder of this chapter our focus will be on which definitions are most beneficial for political and moral discussions regarding capacity-altering biotechnologies.

In many ways our choice of enhancement account might not be all that important. There are multiple ways of defining many terms, and definitions are things that can be stipulated rather than necessarily being correct or incorrect. The account that is chosen, or the definition that is prescribed, merely affects the semantics of the debates. Rather than asking whether memory enhancements (understood as per FA) increase well-being, we could ask whether interventions that increase memory are enhancements (understood as per WC).

However, there are some ways in which the definition of enhancement we use can be important. For example, some definitions may make different sets of questions more salient and easier to respond to. In Section 8.3 we will review some of the ways in which different enhancement definitions can help frame moral debates about capacity-altering biotechnologies. We will focus on the more commonly endorsed enhancement definitions: NFA, BSTA, WC, and FA.

function is understood in a goal-directed way. In an etiological account, we can understand the function of memory to be the recalling of events, as recalling events is what the memory system evolved to do. In this account of function, the diminishment of painful memories would not be considered an enhancement according to FA because it reduces the ability of the memory system to recall events.

[16] Some interventions will be considered enhancements on all seven accounts—that is interventions which increase the functioning of someone who is not diseased, beyond species-typical and -maximum levels, and which increase their chances of having a good life.

Table 8.1 Different enhancement definitions

Approach:	Enhancements are…	Is giving Johnny or Billy growth hormone an enhancement?
Constructivist Approach	beneficial alterations to functioning that do not treat disease, with diseases being understood as states disvalued by society	Johnny but not Billy
Normal Functioning Approach	beneficial alterations to functioning that do not treat disease, with diseases being understood as negative deviations from normal functioning	Johnny but not Billy
Beyond-Species-Typical Approach	alterations that take people beyond species-typical values for particular traits	neither are enhancements
Beyond-Species-MaximumMaximum Approach	alterations that take people beyond species-maximal values for particular traits	neither are enhancements
Welfarist Conception	alterations that improve well-being	both are enhancements
Modified Welfarist Approach	alterations that give people abnormal biological functioning and improve well-being	neither are enhancements
Functional Approach	alterations that increase some type of functioning	both are enhancements

8.3 Framing Enhancement Debates

As stated in the Introduction, some in the bioethics community think the term 'enhancement' is so plagued with problems that we ought not to even use it. What do we gain by conducting debates about the proper use of capacity-altering biotechnologies with the aid of the term 'enhancement'?

When the idea of enhancement was first used in the bioethics literature, it was believed to be useful because it described a group of activities which are inherently morally problematic. In this view the mere fact that the use of a biotechnology would be an enhancement generates moral reasons against it. Take Anderson's view regarding genetic engineering:

> On medical and ethical grounds we should draw a line excluding any form of enhancement engineering. We should not step over the line that delineates treatment from enhancement. (Anderson 1990, p. 24)

Here there is a morally relevant distinction between enhancements and treatment, with the former being universally morally problematic. However, it's not clear that

any definitions of enhancement outlined in Section 8.2 describe something that is universally wrong.

Some theorists believe that the NFA or BSTA definitions of enhancement draw a morally relevant distinction between enhancements and other uses of biotechnology (which are often referred to as 'treatments' on either approach[17]). However, the distinction these definitions draw between enhancements and treatments often seems insignificant. This is because both approaches use arbitrary statistical cut-off points to determine if an intervention is an enhancement or a treatment. Take the following hypothetical example. Let's stipulate that a level of functioning two standard deviations below the mean is pathological, and that the average human IQ is 100 and has a standard deviation of 10. Imagine two people: Jim who has an IQ of 79 and Jane who has an IQ of 81. Both can have their IQs raised to 100 by taking a pill. In NFA, Jim taking the pill is considered to be a therapy and Jane taking the pill is considered to be an enhancement. If the distinction between enhancement and treatment drawn by NFA is necessarily morally significant, it implies that there is a morally significant difference between giving Jane and Jim the pill. Most would find this counterintuitive. An initial difference of two IQ points seems like an insignificant basis for a morally important distinction.

According to BSTA neither Jim nor Jane has been enhanced because their IQs have not exceeded species-typical levels. But suppose that rather than the pill giving both an IQ of 100 it raises Jim's to 119 and Jane's to 121. If we stipulate that species-typical levels are within two standard deviations of the mean, Jane has been enhanced but Jim has not. If the distinction between therapy and enhancement drawn by BSTA is morally significant, it implies that there is a morally significant difference between giving Jane and Jim the pill. Again, this seems counterintuitive. Both have taken a pill which raised their IQ by 40 points, and they now both have IQs that are very similar. There does not seem to be a morally important distinction between their acts.

The fact that the difference between enhancement and treatment often seems arbitrary in this sense has been taken as a reason to reject accounts of enhancement that draw a treatment/enhancement distinction and/or as a reason to deny that such a distinction is morally significant (Harris 2007).

However, even if the treatment/enhancement distinction is often arbitrary, the definitions of enhancement given by NFA and BSTA may still be useful. This is because they draw our attention to morally relevant spectra. For example, BSTA draws distinctions in the spectrum of functioning. This is a continuous spectrum, ranging from very low levels of functioning to very high levels, with species-typical functioning being somewhere in the middle. Similarly, NFA draws distinctions regarding the spectrum of health, with the bodily state of perfect health at one end and the state of

[17] Strictly speaking, in BSTA, enhancement should not be contrasted with treatment. In this approach, many enhancements will still be treatments (as they raise people who have diseases above species-typical values). However, many theorists who use BSTA (e.g. as Harris below) speak of a distinction between enhancement and treatment in this view. We will follow this convention for simplicity.

extreme disease at the other. The main problems with NFA and BSTA may be that they conceive of enhancement in a binary fashion, rather than as a matter of degree (i.e. in a scalar fashion).

Although there may be little difference between cases that are adjacent to each other on these spectra, cases far apart may be very dissimilar. Imagine that in the above case the pill increases both Jim's and Jane's IQ by 40 points, but Jim has a starting IQ of 50 and Jane 110. Now there does seem to be a morally relevant difference between the cases. Say we only have one pill and can only give it to Jim or Jane. Intuitively, and all else being equal, it seems that we have decisive moral reasons to give the pill to Jim. He seems far worse off than Jane, so giving him the pill would be equality-promoting. This points to an important difference between some cases that sit on the treatment/enhancement spectrum. In BSTA (conceived in a scalar rather than a binary fashion), an intervention at the treatment end of the spectrum involves moving individuals, who would otherwise have been below the mean level of functioning, towards the mean. An intervention at the enhancement end of the spectrum will involve moving someone who is already above the mean level of functioning to an even higher level. For traits that are related to well-being and opportunity, this is an important difference. Other things being equal, interventions at the treatment end of the spectrum will tend to promote fundamental equality, whereas interventions at the enhancement end of the spectrum will tend to promote inequality. Similarly, those at the disease end of the spectrum will (other things being equal) tend to be worse off than those at the healthy end. In NFA (conceived in a scalar rather than a binary fashion), an intervention at the treatment end of the spectrum (i.e. aimed at restoring health in someone who is clearly (very) diseased) will tend to be equality-promoting, whereas an intervention at the enhancement end of the spectrum (improving the traits of someone who is clearly (very) healthy) will tend to increase inequality.

Many technologies regularly discussed in the enhancement literature, such as cognitive enhancement and lifespan extension, fit the above picture. Cognitive abilities are linked to opportunities in life. When we move the cognitive abilities of those who are clearly below typical levels towards the mean, we are increasing equality of opportunity. Conversely, when we make someone who has above-average cognitive abilities even smarter, we are increasing their advantages over others and thereby increasing fundamental inequalities. The more the use of a cognitive targeted technology is like a prototypical case of enhancement (i.e. at the enhancement end of the spectrum), the more likely it is to promote inequality. While some may deny that increasing inequality is necessarily, or even often, bad, this suggests that prototypical cognitive enhancement may be considered different from prototypical cases of cognitive therapy. Similar claims apply to life-extension technologies. Allowing those who would have died young to reach typical life expectancies promotes equality. Enabling those who already have long life expectancies to live even longer exacerbates inequality. Therefore the more the use of life-extension technologies is like a treatment (or at the treatment end of relevant spectra), the more likely it is to promote equality, and the more it is like an

enhancement (or at the enhancement end of relevant spectra), the more likely it is to increase inequality. While this does not show that life extension is necessarily wrong, it provides a reason to prioritize therapies above it.

In practical terms the problem with NFA and BSTA is that they are often used to make a binary (treatment/enhancement) distinction across a continuous spectrum. As with other cases of continuous spectra, we should speak in terms of degree and prototypical cases. We should not ask whether something is an enhancement or a treatment, but where particular interventions sit on the spectrum and where they sit in relation to other cases (Selgelid 2014). Often, but not always,[18] where a capacity-altering intervention sits on these spectra will be morally relevant.

The functional and normative accounts of enhancements frame debates about capacity-altering biotechnologies in different ways. WC defines enhancement in terms of something that is clearly good—increasing an individual's chance of living a good life. When defined in this way, the question of whether enhancements are good or bad in general is moot. Enhancements, on WC, are clearly good in some sense. This account of enhancement therefore shifts discussions from focusing on whether we have moral reasons to use enhancements in general, to focusing on which particular biotechnologies should be considered as enhancements. For example, rather than asking if there are moral reasons to use cognitive enhancements, we should ask whether technologies that increase our cognitive capacities are always enhancements. Normative conceptions of enhancement help reframe debates about capacity-altering biotechnologies to focus on whether biotechnologies help us live better lives.

Relatedly, FA does not draw any distinction between enhancements and other uses of biotechnologies which improve functioning. All uses of biotechnology which improve functioning, including treatments, are enhancements on this view. As such, it is clear that the question of whether enhancements are good or bad will depend on the context in which they are used. Few believe that improving the respiratory functioning of someone who is suffering from lung disease would be a bad thing. Radically increasing someone's ability to feel pain, on the other hand, would widely be considered to be bad. In this way, using a functional account of enhancement makes different questions salient. The key questions become: what properties ought we to enhance (or diminish), and what goals should we be pursuing through the use of biotechnologies that alter our traits?

In sum, the term 'enhancement' is still useful for framing modern debates about capacity-altering biotechnologies. While there are many competing definitions of enhancements, each has its own benefits and disadvantages. Those given by NFA and BSTA are useful because they draw attention to morally relevant spectra. Framing debates in these terms helps draw our attention to uses of technologies which tend to

[18] Another oft-discussed case of enhancement in the literature, moral enhancement is unlikely to fit the above picture (Douglas 2008). While moral enhancement may increase inequalities in moral capacities, it may decrease more fundamental inequalities, such as those in well-being and opportunity. This is because they may make individuals more likely to sacrifice their own well-being for others.

increase fundamental inequalities. Other definitions of enhancement, such as FA and WC, help make different questions salient, such as which technologies will increase our chances of living a good life, and which functioning we should increase or diminish. We should therefore resist moves to abandon the term 'enhancement' altogether and instead recognize that there are multiple coherent and useful definitions of enhancement.

8.4 Conclusion

Enhancement debates have long been held back by the lack of a clear consensus about the definition of enhancement. We have identified at least seven different definitions that can be derived from the bioethics literature. Some have suggested that given this lack of consensus we should abandon the term altogether. However, we believe this would be throwing the baby out with the bathwater. The term is still useful for modern debates about capacity-altering biotechnologies. Revisionist notions of enhancement, such as normative and functional accounts, are helpful for reorientating debates about biotechnologies and making particular questions salient, such as which traits we should be increasing or diminishing and which biotechnologies will improve our lives. Traditional notions of enhancement, which define it in terms of altering capacities outside the context of treating disease, or interventions that take people beyond species-typical functioning, can be useful for drawing our attention to morally relevant spectra. Although there is no necessary relationship between something being an enhancement and there being moral reasons against it on these accounts, such definitions can help point us towards contingently morally relevant features.

The most important thing is for theorists to be clear about which definition they are using when they discuss human enhancement. This will prevent meaningful debates about capacity-altering biotechnologies being held back by 'erroneous assumptions' and 'poorly articulated notions of enhancement'.

References

Agar, N. (2010). *Humanity's End: Why We Should Reject Radical Enhancement*. Cambridge, Massachusetts MIT Press.

Allen, D. and Fost, N. (1990). Growth Hormone Therapy for Short Stature: Panacea or Pandora's Box?, *The Journal of Pediatrics* 117(1): 16–21.

Anderson, W. F. (1990). 'Genetics and Human Malleability', *The Hastings Center Report* 20: 21–4.

Boorse, C. (1977). 'Health as a Theoretical Concept', *Philosophy of Science* 44: 542–73.

Canton, J. (2003). 'Convergence for Improving Human Performance', in C. Roco and W. Bainbridge (eds), *Nanotechnology, Biotechnology, Information Technology and Cognitive Science*. London: Kluwer Academic Publishers, pp. 71–9.

Cartwright, S. (1851). 'Diseases and Peculiarities of the Negro Race', *DeBow's Review* 11: 331–6.

Daniels, N. (1994). 'The Genome Project, Individual Differences, and Just Health Care', in F. Murphy and M. Lappé (eds), *Justice and the Human Genome Project*. Berkeley: University of California, pp. 110–32.

Douglas, T. (2008). 'Moral Enhancement', *Journal of Applied Philosophy* 25: 228–45.

Earp, B., Sandberg, A., Kahane, G., and Savulescu, J. (2014). 'When is Diminishment a Form of Enhancement? Rethinking the Enhancement Debate in Biomedical Ethics', *Frontiers in Systems Neuroscience* 8(12): 1–7.

Elliott, C. (2003). *Better Than Well: American Medicine Meets the American Dream*. New York: W.W. Norton.

Engelhardt, H. T., Jr (1974). 'The Disease of Masturbation: Values and the Concept of Disease', *Bulletin of the History of Medicine* 48: 234–48.

Erk, C. (2011). *Health, Rights and Dignity*. Berlin: De Gruyter.

Harris, J. (2007). *Enhancing Evolution*. Princeton, NJ: Princeton University Press.

Juengst, E. (1998). 'What Does *Enhancement* Mean?', in E. Parents (ed.), *Enhancing Human Traits: Ethical and Social Implications*. Washington, DC: Georgetown University Press, pp. 29–47.

Kingma, E. (2007). What Is it to Be Healthy?', *Analysis* 67: 128–33.

Little, W., Fowler, H., and Coulson, W. (eds) (1959). *The Shorter Oxford English Dictionary*. Oxford: Oxford at the Clarendon Press, p. 612.

Murphy, D. (2015). 'Concepts of Disease and Health', in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, spring edn, <http://plato.stanford.edu/archives/spr2015/entries/health-disease/>, accessed 18 November 2015.

Parens, E. (1998). 'Is Better Always Good?: The Enhancement Project', *The Hastings Center Report* (special supplement) 28: s1–s17.

President's Council on Bioethics (ed.) (2003). *Beyond Therapy: Biotechnology and the Pursuit of Happiness*. Washington DC: US Government Printing Office.

Resnik, D. (2000). 'The Moral Significance of the Therapy-Enhancement Distinction in Human Genetics', *Cambridge Quarterly of Healthcare Ethics* 9: 365–77.

Savulescu, J., Sandberg, A., and Kahane, G. (2011). 'Well-Being and Enhancement', in J. Savulescu, R. ter Meulen, and G. Kahane (eds), *Enhancing Human Capacities*. Chichester: Blackwell Publishing, pp. 3–18.

Schwartz, P. H. (2005). 'Defending the Distinction between Treatment and Enhancement', *The American Journal of Bioethics* 5(3): 17–19.

Selgelid, M. J. (2014). 'Moderate Eugenics and Human Enhancement', *Medicine, Health Care and Philosophy* 17: 3–12.

Wiley, E. (1978). 'The Evolutionary Species Concept Reconsidered', *Systematic Biology* 27: 17–26.

9

Human Enhancement for Whom?

Robert Sparrow

Whose interests matter when making decisions about what sort of children to have? Although it has received scant attention, this question is at the heart of the ethics of reproductive decision-making. It is also crucial to the ethics of human enhancement given that the most powerful technologies of human enhancement are likely to involve shaping future individuals via genetic selection or genetic modification. If we wish to have a sensible debate about human enhancement, then, we must first become clear on the question of 'enhancement for whom?' (Elster 2011). In this chapter I survey and evaluate the claims of the three leading candidates whose interests might be thought to matter when it comes to shaping future people: the parents, the child, and 'the world'. I also discuss what I take to be two bad candidates, which nevertheless seem likely to loom large when it comes to popular discussions of this topic—'the race' and 'the species'—as well as one candidate that is more plausible, although still, I think, properly controversial—the nation. I will argue that the parents, child, and 'world' all have legitimate interests in reproductive decisions and that these interests may conflict more than has been appreciated. For this reason, enhancement is more ethically problematic than proponents typically admit. The danger that policy on enhancement will in practice be guided by concern for the interests of the nation or, worse, the race constitutes a further reason for caution about the 'enhancement enterprise'.

9.1 Reproduction, Enhancement, and the Interests at Stake

Couples considering having children must confront a number of decisions. First, a couple must decide *whether* or not to have children. Second, they must decide *when* to have children. In making this decision couples also determine in a small way the answer to a third question: '*What sort of* children should they have?' By choosing to delay having children until they are older and more financially secure, for example, couples influence the expected welfare of their child. The development of new reproductive technologies, and in particular the technology of preimplantation genetic diagnosis (PGD), poses this third question with a new urgency.

PGD allows prospective parents some control over the genetics of their future child by allowing them to choose which of a number of embryos, created via IVF, to implant into a woman's womb on the basis of a knowledge of the genetics of each embryo. This is a powerful technology for selecting against genes associated with particular phenotypes. Inevitably it is also a (weak) technology for selecting in favour of genes associated with desired phenotypes. The usefulness of PGD for positive selection is currently limited both by our understanding of the relationship between genotypes and phenotypes and by the small number of embryos that can be created in each cycle of IVF. However, improvements in computing and sequencing technology hold out some prospect of addressing the former of these problems. The development of a technology to produce ova from somatic cells via induced pluripotent stem cells ('artificial gametogenesis') may overcome the latter. If parents were able to choose from among thousands of embryos created using artificial gametes, the utility of PGD for positive selection would be greatly increased (Bourne, Douglas, and Savulescu 2012; Sparrow 2013; Palacios-González, Harris, and Testa 2014). Other, more speculative, technologies involving the genetic modification of embryos (discussed further in Section 9.1.2) may grant parents still greater power over the genetics of their children.

Thus, as technology provides us with the power to determine the genetics of our children, the question of what sort of children we should have arises with increased urgency. This question is central to the ethics of human enhancement because one of the most plausible ways to enhance human beings would be to alter them genetically.

However, before we can answer *this* question there is another matter we first must become clearer about: we need to know what makes one sort of child better than another. What makes something an 'enhancement'? It might be thought that enhancements are whatever make something better and that we are simply enhancing 'capacities'. Yet while these answers may do in some contexts, at a deeper level they fail: ultimately we need to know *why* one state of affairs counts as better than another when we are thinking about the ethics of reproduction. Presuming that what is better is better for someone, we need to know whose interests are relevant when it comes to decisions about what sort of children to have. For whom are alterations to the genetics of children an enhancement? In the following sections I survey and evaluate a number of different answers that might be given to this question. I will also offer some thoughts about how we might go out about balancing the competing interests I identify.

9.1.1 The Parents

The obvious place to begin a list of the interests at stake in reproduction is with the parents. A non-trivial portion of the project of having children takes place before the children are born—some of it before they are conceived. If couples have made a conscious choice to reproduce, there will be reasons for this choice, and the desires that motivate these reasons may ultimately be satisfied or frustrated. Even if parents have not made a conscious choice to reproduce, their lives will usually become intertwined

with the lives of their children such that they will come to have an interest in their children's flourishing.

There is one particular interest that parents have that is worth singling out for attention because of its centrality to contemporary debates about the ethics of reproductive technologies—an interest in the children whom they bring into existence being genetically related to them. Not all parents—or prospective parents—care about this, and many parents realize their interests in reproduction without the child whom they raise being the product of their gametes. Nevertheless, many individuals desire not only to bring children into existence but for these children to be their genetic offspring. Indeed, so much is this the case that in many contexts genetic parenthood is assumed to be the very definition of parenthood. As we shall see below, in Section 9.1.2, however, insisting on this particular form of parenthood may sometimes be at the expense of the interests of other parties implicated in reproductive decisions.

The moral weight of the interests of the parents is reflected in the intuitive and legal force of the right to 'reproductive liberty' (or 'procreative autonomy'). The place of reproduction at the heart of (many) ideas about what a good human life consists in establishes a strong presumption against coercive interference in couples' reproductive projects (Dworkin 1993; Brock 1994; Robertson 1994, especially pp. 22–42; Savulescu 1999; Buchanan et al. 2000, pp. 206–13). While the precise extent of the activities that are properly understood to be defended by this right is contested (Sparrow 2008; de Melo-Martín 2013), at the very least prospective parents have a right against forced sterilization or abortion, and against being forced to become parents against their will by (for instance) being raped or denied access to contraception.

However, as the debate about the *limits* of reproductive liberty makes clear, there are other interests at stake in decisions about reproduction (Brock 1994). Most obviously, the interests of the parents may conflict with the interests of the child.

9.1.2 The Child

By altering the timing of conception (Parfit 1984, pp. 357–61, 367–9), or by choosing different embryos via PGD, parents may bring into existence a child with a severe disorder or a one that is perfectly healthy. The fact that parental decisions have such an impact on the lives of their children means that those children clearly have interests at stake.

The interests of the child in reproductive decisions have been championed by Julian Savulescu (Savulescu 2001; Savulescu and Kahane 2009). He points out that the same reasons that move parents to embrace genetic selection for therapeutic purposes—a concern for the welfare of the future child—should also move them to embrace selection for beyond-species-typical traits (enhancement) where possible. This emphasis on the interests of the child serves to distinguish contemporary enthusiasm for human enhancement from historical programmes of eugenics, which were oriented towards the interests of the race or nation (Savulescu 2001, p. 424).

The moral weight of a concern for the interests of the child arguably depends on the nature of the reproductive technology we are considering. Decisions made before the child comes into existence seem to have a different moral import to decisions made once they exist (Parfit 1984). Because it is possible to identify the impact of decisions made any time after conception on the welfare of an individual, most authorities have tended to treat modifications to the embryo, or interventions which affect the developing foetus, as harming or benefiting particular individuals. As a result it will usually be morally problematic to sacrifice the welfare of children without their consent for the sake of the interests of third parties.

However, as Derek Parfit famously argued, decisions that determine *which* individuals come into existence do not seem to harm or benefit those individuals (Parfit 1984). The counterfactual claim about what their interests would have been had the decision been made differently, which is usually thought to be essential to the determination of harm or benefit, fails in such cases because had the decision been made differently a different person would have been brought into existence. Such decisions are not 'person-affecting', and because of this their ethics is both controversial and peculiar. For instance, it is this aspect of the ethics of genetic selection that explains one of the distinctive features of Savulescu and Kahane's putative obligation of procreative beneficence, which is that it is—as they themselves readily admit—so easily defeasible (Savulescu and Sparrow 2013).[1]

Until recently the most realistic hope for enhancing human capacities by shaping individuals' genes involved genetic selection for beyond-species-typical traits, perhaps over many generations or in conjunction with the use of artificial gametogenesis (Sparrow 2013). Such genetic selection would not be person-affecting. I will mostly be concerned with choices of this sort here. However, the development of the CRISPR/Cas9 system for genome editing (Cong et al. 2013) has opened up the possibility that we might be able to perform very precise genetic manipulations of human embryos in the not too distant future, including the introduction of genes for the purposes of enhancement (Regalado 2015). Whether such a technology would be person-affecting or not is unclear. If we think of it as involving the manipulation of an embryo that we have already chosen to come into existence, and think of the changes we make as benefiting a particular individual, then it will be person-affecting. However, if we think of it as requiring a choice about *which* person to bring into existence, either because the technology might also involve some selection *between* embryos or because we think of

[1] Savulescu and Kahane (2009) argue that procreative beneficence provides parents with a 'significant moral reason' to select the best child possible. However, they leave both the notion of a 'moral reason' and the question of the weight of 'significant' moral reasons underdescribed. Although they claim that significant reasons are 'often strong enough to outweigh the reasons given by the interests of parents and other existing people' (p. 277), they also explicitly allow that the political commitments of parents or concerns about aggregate consequences may justify parents ignoring the demands of procreative beneficence (Savulescu 2001; Kahane and Savulescu 2010). Given this it is hard to see why, on their account, anything the parents want strongly enough would not also justify the ignoring of procreative beneficence.

(some) alterations of genetic traits as themselves being identity-affecting (Zohar 1991), then this will not be a person-affecting technology.

Another key issue in the debate about the moral significance of the interests of the child concerns the extent to which these interests are amenable to comparison and (therefore) to maximization. If it were not at least sometimes the case that a child's genetic endowment at birth had implications for their expected welfare, then the interests of the child could play no role in determining the ethics of genetic selection or modification. Although the claim has been vigorously contested by writers within the disabilities community (Hurst 2009; Garland 2015; see also Barnes 2014), many people share the intuition that, for instance, having the gene for cystic fibrosis reduces the expected level of welfare of a child such that all other things being equal it is better to be born without this gene (Glover 2006; Harris 2007; Savulescu and Kahane 2009). Such judgements about the impact of genes associated with various impairments on expected welfare are central to the ethics of genetic selection for therapeutic purposes. If there is to be an obligation to enhance human beings by selecting for genes—or perhaps introducing genes—for capacities beyond those that are species-typical, then it must also be the case that those born with these genes would have higher expected welfare than those born without them.

The demands of the argument for human enhancement would be strongest were it the case that it is always possible to compare the expected welfare of children at birth and determine which of them had the highest expected welfare. I suspect that many people interpret Savulescu's claim that parents are obligated to have the best child possible—as I did in earlier work (Sparrow 2007, 2011a)—as implying that there was always an answer to the question as to which of two (or more) infants had superior life prospects.

However, more recently, Savulescu has conceded that in many cases there will simply be no answer to the question as to which of two children has better life prospects at birth (Savulescu 2015). One child may have a gene that makes it easier for them to succeed as a musician, while another has a gene that makes it easier to succeed as a poet, and there may be no answer as to whether it is better to be a musician or a poet.

While this seems eminently plausible, it calls into question the force of the case for human enhancement. The proper formulation of the obligation of procreative beneficence is revealed as 'the obligation not to have a child with life prospects worse than any other child one might have bought into existence' (Sparrow 2014a). Moreover, given the diversity of ways of flourishing available to any person born with even a relatively meagre set of characteristically human capacities, it is likely that the life prospects of almost all embryos should be judged to be 'equally good', and thus that this obligation will usually be moot. Only in the rare case where one embryo had genes that would provide the future child with no less capacity to succeed in any life plan than the genes of another embryo, while also providing better prospects for success in one or more life plans, would it be possible to judge that some genetic endowments were superior to others without the fact of the plurality of worthwhile life plans rendering

evaluations of the relative advantages of different combinations of genes impossible. Moreover, it seems likely that most genes that benefit some life plans will also be to the detriment of others (Sparrow 2010a, 2012a).

There are, of course, many ways in which parents may fail to promote the interests of their children: any decision that does not maximize the capacities or the welfare of the child will be contrary to the child's interests. Two, however, are worth highlighting here because of their role in the larger debate about the ethics of technologically mediated reproduction.

First, parents may act against the interests of the child when they insist on having a child who will be genetically related to them. While PGD may allow parents to choose the best of the embryos produced by the combination of their gametes, in many cases all of their embryos will have poorer life prospects than an embryo created using the gametes of third parties.[2] Thus an obligation to maximize the welfare of their children would often require parents to use donor gametes (Sparrow 2007, 2011a). Savulescu and Kahane (2009) later qualified the 'obligation' of procreative beneficence to apply only to choice among the possible genetic offspring a couple might have. However, this restriction seems unprincipled.

Second, the choices parents might want to make regarding their children's genetic endowment, based on their own ideals and values, may not be in the best interests of the child (Davis 1997). While, at least in the context of genetic selection—with the exception of a preference for using their own gametes—it will (hopefully) be rare for parents to consciously sacrifice the interests of their children for their own interests, what seems more likely is that parents will choose genes that will help the child realize the ends that the parents believe to be worthwhile. Unless the genes the parents choose will make it no harder for children to pursue any (and every) life plan they might come to endorse, there will be a chance that, if the children do not grow up to share their parents' ends, they may find their goals frustrated by their parents' choices (Agar 1999). Again, as I suggested earlier in this section, many genes—perhaps even the vast majority—will benefit individuals in the pursuit of some goals but be deleterious if they have others. What is at stake here, then, is the child's 'right to an open future' (Feinberg 1980; Buchanan et al. 2000, pp. 170–2). Yet the question of which futures are more open cannot ultimately be resolved without reference to ideas about which ends are worth pursuing (Sparrow 2010a). A concern for the best interests of the child may therefore either require us to settle on an account of the nature of human flourishing, which will inevitably be controversial, or prohibit us from bringing children into existence with anything other than 'general-purpose' genes (Fox 2007; Dekker 2009; Sparrow 2010a; Fowler 2015).

[2] Setting aside for one moment the possibility that most such judgements fail in the face of incommensurability, in which case, as I observed earlier, arguments based on the interests of the child will, for the most part, be moot.

While recent discussions about the ethics of human enhancement have been dominated by concern for the welfare of the child, it did not take long for ideas about collective interests to re-emerge. As we shall see, however, the relevant collective remains contested.

9.1.3 The 'World'

The suggestion that we should be concerned with the interests of people other than the parents who are reproducing or the children they might bring into existence appears in the debate about the ethics of genetic selection whenever people refer to the aggregate consequences of reproductive choices. For instance, the objection that it would be a bad thing if the widespread use of PGD for the purposes of either therapy or enhancement led to a loss of 'human diversity' relies on the intuition that such diversity makes the world a better place (Parens 1995; Garland-Thomson 2012, 2015; Sparrow 2015).

The idea that a larger community may have an interest in reproductive decisions has also been defended more systematically. In a wonderful piece of applied philosophy, Jakob Elster (2011) points out that the arguments that Savulescu uses to motivate the 'obligation' of procreative beneficence also imply an obligation to promote the interests of parties other than the child.[3] Just as it is irrational not to prefer a child with higher expected welfare over a child with lower expected welfare, so too is it irrational—all other things being equal—to prefer a child who would contribute less to the welfare of the world one who would contribute more. Just as common-sense morality recognizes that we should prefer our children to be born healthy and happy, so too does it recognize that we should prefer our children to make a contribution to the happiness of others. Douglas and Devolder (2013) have argued for the existence of what they call an obligation of 'procreative altruism' along similar lines.

Once one begins to consider one's obligation to third parties, there is no obvious limit as to how far these obligations might extend. Procreative altruism therefore requires us to consider the interests of 'the world' when making reproductive decisions. The welfare of the world is, in turn, the aggregate welfare of all the creatures with interests in the world.[4]

Importantly, there is another route whereby one might arrive at the conclusion that parents are obligated to promote aggregate welfare in their reproductive decisions. As we saw in Section 9.1.2 above, decisions about which individuals to bring into existence appear to neither harm nor benefit the individuals we do bring into existence. This fact makes it difficult to theorize our obligations in relation to these choices: given that they don't harm or benefit anyone it might be argued that they are morally equivalent. Nevertheless, many people do have the intuition that—all other things being equal—it would be wrong to bring an individual into existence with low expected welfare when we could have bought an individual with a higher expected welfare into

[3] See also Petersen 2015.

[4] It is a further question just what this set consists in.

existence instead. Such an obligation requires us to care about benefits (or harms) that are not benefits (or harms) to anyone—that is, about our impact on 'impersonal' welfare (Parfit 1984, pp. 386–7; Bennett 2009).

Yet once we take impersonal welfare into account, it seems implausible that we should not also care about the impact of our decisions on the welfare of existing individuals. Thus, again, it appears as though when making reproductive decisions we should consider the interests of all the individuals who exist, or will exist, regardless of our decision, as well as those who will exist as a result of our decision.

There is a large and complex philosophical literature that explores the implications of a concern for impersonal welfare for the ethics of reproductive decision-making. Notoriously, a concern for aggregate welfare that includes impersonal welfare seems to require parents to have as many children as possible. Moreover, the various argumentative tweaks that philosophers have attempted to try to avoid this implication all seem to generate equally unappealing consequences. Reasons of space prevent me from discussing this literature and its complexities here. Instead I must be content to make two observations.

First, for the most part, this literature has neglected the options established by the possibility of genetic selection for enhancement for the promotion of aggregate welfare. Presumably, for instance, it would improve aggregate welfare if a small number of parents were to bring into existence children who had been cloned and/or genetically modified in order to be anencephalic for the purposes of serving as sources of replacement organs for other citizens (Tooley 1998). As I have argued elsewhere, it seems likely that replacing a very small number of the members of a population with uniformly high welfare with individuals who will have lives only barely worth living would increase aggregate welfare by providing the individuals with higher welfare with an opportunity to be happy that they are better off than these 'genetic scapegoats' (Sparrow 2015). Other modifications, such as increasing the extent of human altruism (Persson and Savulescu 2012) or individuals' willingness to take on various necessary social roles, such as collecting the garbage or caring for the elderly, might also be expected to make a significant contribution to increasing aggregate welfare.

Second, the moment one admits any role for considerations of aggregate welfare in decisions about reproduction, these threaten to outweigh any of the other interests at stake. Given the number of third parties who might benefit from the selection of future individuals that have particular sorts of capacities, there is little that the concern for the interests of such third parties could not justify.

9.1.4 The Species

Arguments about the implication of genetic technologies for the 'species' come up surprisingly often in both the literature and public discussions of this topic. Thus, for instance, when I have suggested that a concern for the best interests of the child—procreative beneficence—would require all parents to choose female children on the basis of their longer life expectancy and more 'open' future (Sparrow 2010b, 2012b),

critics have often responded that this would be disastrous for the species (Harris 2011; Casal 2013). Similarly, people often worry about the impact of the widespread use of technologies of genetic selection on the extent of the genetic diversity within our species, and on the likelihood that the species might suffer extinction as a result of some future pandemic or environmental change (Suzuki and Knudtson 1989; Bayliss and Robert 2004, pp. 7–8). Conversely, Chris Gyngell (2012) and Russell Powell (2012) have argued that genetic human enhancement might contribute to the long-term survival of the human species.

Such pleading on behalf of the species is surprising because it is far from obvious why the fate of the species should be of concern to us. Species do not experience anything, nor do they suffer or die, except metaphorically (Singer 1979, p. 203). The welfare of the individual organisms that make up a species matters, of course, but events or policies may lead to a species becoming extinct without harming any of its members. For instance, the aggregate welfare of the world might be vastly improved were all children from the next generation onwards to be born 'post'-human with greatly improved capacities and (consequently) no interest in reproducing with members of previous generations, who they view as pathetically short-lived and profoundly cognitively impaired. In such circumstances the human species would become extinct without any individual suffering.

Even if species do have some value, this fails to establish why the survival of *Homo sapiens* should be of any more concern to us than the survival of any other species. Sadly, it seems likely that the extinction of the human species would actually greatly *reduce* the number of species that are likely to become extinct in the next several decades as a result of humankind's environmentally destructive activities, which are currently estimated to be driving one in six species to extinction (Urban 2015). Concern for the value of species, then, provides us with little reason to regret the extinction of the human species.[5]

9.1.5 The Nation

Because the implications of genetic selection for the world or the species are a function of the aggregate consequences of millions of reproductive decisions, it is actually extremely difficult to take account of them when assessing either the ethics or the rationality of individual decisions. Given that any individual decision will make only an infinitesimal contribution to the aggregate impact, even a small reason for making the choice that—if universalized—would generate the undesirable aggregate impact will typically outweigh concerns about the aggregate impact. For instance, even if parents recognize that a world in which everyone were born female would be a worse world, they may still have reasons to choose a female child themselves as long as they think that a girl will have better life prospects no matter what decisions other parents

[5] Clearly some people do care about the future of the human species, so a concern for their welfare may also dictate a concern for the welfare of the species.

make (Sparrow 2011b). Avoiding such consequences therefore involves solving a collective action problem. Collective action problems are inevitable where what parents are pursuing for themselves or their child is a positional good.

The best way to resolve collective action problems is for all those involved to cede (some) power to an institution that is capable of reshaping the option sets of individual decision-makers by imposing penalties on anyone who chooses to 'defect' from the course of action required to secure a collective benefit. In practice, this typically means a national government. While, in theory, national governments could legislate to defend the interests of *Homo sapiens* or to promote aggregate welfare at a global level, historically governments have tended to legislate in the national interest. Thus, for instance, when states—as they often do—adopt population policies, the justification for them usually refers to sex ratios or birth rates nationally rather than globally. Moreover, the transnational institutions required in order to regulate human enhancement in the interests of the world neither exist nor look very appealing once we start to imagine what sorts of powers they would require.

To avoid the adverse aggregate consequences of parents each pursuing either the best interests of their child *or* their own best interests through genetic selection, it will be necessary to regulate access to, and the uses of, technologies of genetic selection. Such regulation is likely to be directed by a concern for the national interest. As I will discuss further in Section 9.2 below, human enhancement in the national interest might require significant sacrifices from both parents and children.

9.1.6 The Race

The largest ever program of 'human enhancement' to date was actually oriented towards the enhancement of a 'race' rather than a population understood as coextensive with the boundaries of a state, as the nation is more commonly thought of today. Both the National Socialist programmes of positive eugenics, through the SS's Marriage Decree of 1931 and the *Lebensborn* programme, and the negative eugenics of the 1933 Act for Averting Descendants Afflicted with Hereditary Disease, which allowed for compulsory sterilization, were intended to promote the future welfare of the 'Aryan race' (Hubbard 2010). Similarly, the eugenic policies adopted by other nations around the world between the 1880s and 1940s were typically concerned with the welfare of populations conceived of through racialized narratives and were justified by reference to racial stereotypes (Kevles 1999).

It would be comforting to believe that the history of eugenics has irredeemably tarnished the idea that reproductive decisions should be made in the interests of the race. Unfortunately, however, there are a number of reasons to worry about the re-emergence of a racialized eugenics in the context of the enhancement enterprise (Sparrow 2014b). As I argued in Section 9.1.5, human enhancement—if it is regulated at all—is likely to be regulated with reference to the national interest. The genetics of the nation's population is likely to be a central concern of such regulation. However, such 'geneticization' of the idea of the nation is fraught with political risks. There are

significant tendencies within nationalism that have traditionally sought to represent the nation as the political expression of the interests of a people, and the people as racially distinct from other peoples. Placing the genetics of the nation's population at the heart of regulations concerning a choice as ubiquitous and as intimate as reproduction seems likely to lend strength to this tradition. Moreover, once such a project is under way, there will be a temptation to allow eugenic arguments to play a role in other decisions that impact on the population's genetics, including decisions about healthcare and immigration. Not only would this further entrench the idea that all citizens share a common genetic heritage but also it would inevitably encourage the idea that non-nationals are genetically distinct and are a threat to the 'good' genetics of the nation. Given that we ordinarily act in more or less complete ignorance of the actual genes of other people, in practice such suspicion will naturally tend to adhere to those who look 'foreign' or 'different' in some way, which is to say that they will both track and reinforce historical ideas about 'race'. For all these reasons there is, I believe, a significant danger that the regulation of human enhancement in the national interest will lead to the reinvigoration of a 'scientific' racism.

9.2 A Difficult Balancing Act?

Any plausible programme of genetic human enhancement would need to come to some conclusion about how to balance the different interests that—as I have argued here—are, or might be thought to be, at stake in reproduction. I believe that this task is much more difficult than generally appreciated and also, consequently, that the project of genetic human enhancement is more ethically fraught than proponents generally admit.

If the ethics of reproduction were simply a matter of balancing the interests of the parent against the interests of the child then this dilemma might at least be manageable, even if the fact that decisions about which children to bring into existence are not person-affecting means that it is more complex than first appears. However, the introduction of even the very basic reproductive technologies of contraception, selective abortion, and sex selection are sufficient to establish the possibility of collective action problems, which requires the state to step in and legislate. This in turn requires us to try to balance the reproductive liberty of parents against the interests of the larger collective, and the interests of the child against both.

The development of genetic technologies that might allow us to shape the capacities and character of our children would greatly intensify these challenges. As we have seen, it would increase the likelihood of parents acting against the interests of their children when they shape them in ways that the children may come to regret, which strengthens the case for the state to intervene to protect the interests of children. This would inevitably involve the state acting to adjudicate over what sorts of genetic intervention are incompatible with the child's right to an open future, a role which is itself incompatible with the 'state neutrality' that many liberals have understood as essential

to a just society (Dekker 2009; Sparrow 2010a). Granting parents the ability to consciously shape the genetics of their children would also greatly increase the likelihood that parental choices would establish collective action problems, either because parents would be motivated to seek positional goods for themselves or their children, or because the aggregate consequence of parents all seeking the same sorts of goods for themselves or their children would involve harms to the world, nation, or species. Again, in order to mitigate these risks, the state would need to regulate human enhancement and constrain the reproductive liberty of parents. Finally, as it becomes more powerful, genetic enhancement will greatly increase the extent to which it is possible to engineer human beings for the benefit of the world, nation, or species.

This last possibility is especially disturbing. While advocates of procreative altruism have written of the need to 'balance' the interests of the world against the other interests at stake in decisions about genetic human enhancement, to my knowledge no one has yet described or defended any principle that might plausibly guide us in this task. There are at least two reasons to believe this will be very difficult. First, once one admits any role for the interests of third parties, this opens up the possibility that these interests might be very significant. For instance, an appropriately ambitious programme of society-wide genetic engineering—the creation of a 'Brave New World'—might hold out the promise of a tremendous increase in aggregate welfare. It would require very significant competing interests to balance our obligation to bring about this good. Second, that both the interests of parents and the interests of children may sometimes be outweighed by the collective good must already be conceded by advocates of the regulation required to avoid destructive collective action problems, which will be essential to any defensible 'liberal' programme of genetic human enhancement. Any suggestion that these interests will reliably trump the interests of third parties is therefore grossly implausible. The fact that decisions to bring individuals with lower welfare into existence do not harm those individuals that are brought into existence is a further reason to think that arguments about the interests of children will ultimately prove little barrier to the pursuit of the public good by genetic means.

As I observed in Section 9.1.5, the interests of 'the world' will in practice be attended to by the representatives of the nation, who are likely to be all too quick to adopt the national interest as the guiding principle for the regulation of human enhancement. The answer to the question 'enhancement for whom?', then, will turn out to be enhancement for the nation. Moreover, what is good for the nation may be very bad indeed for our children and for us.

This is, ultimately, why I believe the project of enhancing our children's genes is more ethically problematic than advocates typically acknowledge. People might prefer for no one to have this power over the character of their children than to end up in a situation where they have supported the development of the technology that might enable them to exercise this power themselves only for the state to usurp it and wield it at the expense of their interests and the interests of their children.

However, even if this is an accurate assessment of the conclusion of the 'enhancement enterprise', the thought is unlikely to be sufficient to prevent societies from embarking on it. At least initially, the power that genetic selection or manipulation will grant us over the character of our children is likely to be slight and analogous to the power that parents already have to shape their children through environmental manipulation. Consequently, the sacrifices required of parents and their children by the state in order to regulate enhancement appropriately early on are likely to be small and thus little barrier to public acceptance of the project. Only when the technology to engineer some people for the benefit of others is developed and the argument for a Brave New World made loudly in public again may people realize what they bought into when they bought into genetic human enhancement.

Yet, as discussed in Section 9.1.6, there is another danger associated with the regulation of human enhancement in the national interest, which is likely to be pressing even when technologies of genetic human enhancement remain relatively rudimentary: the danger of reinvigorating the racism that is historically associated with eugenics. While the connection between the project of shaping the genetics of a national population and the re-emergence of a racialized eugenics is a political and contingent one, it is nonetheless strong and should, I believe, motivate us to be very cautious indeed about embarking on the project of human enhancement by genetic means. Perversely, then, if I am right about the dangers ultimately associated with the prospect of genetic human enhancement, critics of this project may have some reason to be thankful that its historical legacy is so shameful.

Acknowledgements

The research for this chapter was supported by the Australian Research Council's Future Fellowships funding scheme (project FT100100481). The views expressed herein are those of the author and are not necessarily those of the Australian Research Council. Thanks are also due to Catherine Mills, Steve Clarke, and Alberto Giubilini for comments on a draft of this chapter, and to Mark Howard, who assisted me with proofreading and with locating various sources while I was writing.

References

Agar, N. (1999). 'Liberal Eugenics', in H. Kuhse and P. Singer (eds), *Bioethics: An Anthology*. Oxford: Blackwell, pp. 171–81.

Barnes, E. (2014). 'Valuing Disability, Causing Disability', *Ethics* 125(1): 88–113.

Bayliss, F. and Robert, J. S. (2004). 'The Inevitability of Genetic Enhancement Technologies', *Bioethics* 18(1): 1–26.

Bennett, R. (2009). 'The Fallacy of the Principle of Procreative Beneficence', *Bioethics* 23(5): 265–73.

Bourne, H., Douglas, T., and Savulescu, J. (2012). 'Procreative Beneficence and *In Vitro* Gametogenesis', *Monash Bioethics Review* 30: 29–48.

Brock, D. (1994). 'Reproductive Freedom: Its Nature, Bases and Limits', in D. Thomasma and J. Monagle (eds), *Health Care Ethics: Critical Issues for Health Professionals*. Gaithersburg: Aspen Publishers, pp. 43–61.

Buchanan, A., Brock, D. W., Daniels, N., and Wikler, D. (2000). *From Chance to Choice*. Cambridge: Cambridge University Press.

Casal, P. (2013). 'Sexual Dimorphism and Human Enhancement', *Journal of Medical Ethics* 39(12): 722–8.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013). 'Multiplex Genome Engineering Using CRISPR/Cas Systems', *Science* 339(6121): 819–23.

Davis, D. S. (1997). 'Genetic Dilemmas and the Child's Right to an Open Future', *Hastings Center Report* 27(2): 7–15.

de Melo-Martín, I. (2013). 'Sex Selection and the Procreative Liberty Framework', *Kennedy Institute of Ethics Journal* 23(1): 1–18.

Dekker, T. J. (2009). 'The Illiberality of Perfectionist Enhancement', *Medicine, Health Care, and Philosophy* 12: 91–8.

Douglas, T. and Devolder, K. (2013). 'Procreative Altruism: beyond Individualism in Reproductive Selection', *Journal of Medicine and Philosophy* 38(4): 400–19.

Dworkin, R. (1993). *Life's Dominion: an Argument about Abortion, Euthanasia, and Individual Freedom*. New York: Knopf.

Elster, J. (2011). 'Procreative Beneficence—Cui Bono?', *Bioethics* 25(9): 482–8.

Feinberg, J. (1980). 'The Child's Right to an Open Future, in W. Aiken and H. LaFollette (eds), *Whose Child? Children's Rights, Parental Authority, and State Power*. Totowa, NJ: Littlefield, Adams & Co., pp. 124–53.

Fowler, T. (2015). 'In Defence of State Directed Enhancement', *Journal of Applied Philosophy* 32(1): 67–81.

Fox, D. (2007). 'The Illiberality of "Liberal Eugenics"', *Ratio* 20(1): 1–25.

Garland-Thomson, R. (2012). 'The Case for Conserving Disability', *Journal of Bioethical Inquiry* 9(3): 339–55.

Garland-Thomson, R. (2015). 'Human Biodiversity Conservation: a Consensual Ethical Principle', *The American Journal of Bioethics* 15(6): 13–15.

Glover, J. (2006). *Choosing Children: Genes, Disability, and Design*. Oxford: Oxford University Press.

Gyngell, C. (2012). 'Enhancing the Species: Genetic Engineering Technologies and Human Persistence', *Philosophy & Technology* 25(4): 495–512.

Harris, J. (2007). *Enhancing Evolution: the Ethical Case for Making Better People*. Princeton, NJ: Princeton University Press.

Harris, J. (2011). 'Sparrows, Hedgehogs and Castrati: Reflections on Gender and Enhancement', *Journal of Medical Ethics* 37(5): 262–6.

Hubbard, R. (2010). 'Abortion and Disability: Who Should and Should not Inhabit the World?', in L. J. Davis (ed.), *The Disability Studies Reader*, 3rd edn. New York and London: Routledge, pp. 107–19.

Hurst, R. (2009). 'Are Disabled People Human?', in P. Healey and S. Rayner (eds), *Unnatural Selection: the Challenges of Engineering Tomorrow's People*. London and Sterling, VA: Earthscan, pp. 60–6.

Kahane, G. and Savulescu, J. (2010). 'The Value of Sex in Procreative Reasons', *The American Journal of Bioethics* 10(7): 22–4.

Kevles, D. J. (1999). 'Eugenics and Human Rights', *British Medical Journal* 319(7207): 435–8.

Palacios-González, C., Harris, J., and Testa, G. (2014). 'Multiplex Parenting: IVG and the Generations to Come', *Journal of Medical Ethics* 40(11): 752–8.

Parens, E. (1995). 'The Goodness of Fragility: on the Prospect of Genetic Technologies Aimed at the Enhancement of Human Capabilities', *Kennedy Institute of Ethics Journal* 5(2): 141–53.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Clarendon Press.

Persson, I. and Savulescu, J. (2012). *Unfit for the Future: the Need for Moral Enhancement*. Oxford: Oxford University Press.

Petersen, T. S. (2015). 'On the Partiality of Procreative Beneficence: a Critical Note', *Journal of Medical Ethics,* 23 April, doi: 10.1136/medethics-2014-102420.

Powell, R. (2012). 'The Evolutionary Biological Implications of Human Genetic Engineering', *Journal of Medicine and Philosophy* 37(3): 204–25.

Regalado, A. (2015). 'Engineering the Perfect Baby', *MIT Technology Review*, 5 March, <http://www.technologyreview.com/featuredstory/535661/engineering-the-perfect-baby/>.

Robertson, J. (1994). *Children of Choice: Freedom and the New Reproductive Technologies*. Princeton, NJ: Princeton University Press.

Savulescu, J. (1999). 'Sex Selection: the Case For', *Medical Journal of Australia* 171(7): 373–75.

Savulescu, J. (2001). 'Procreative Beneficence: Why We Should Select the Best Children', *Bioethics* 15(5–6): 413–26.

Savulescu, J. (2015). 'Procreative Beneficence, Diversity, Intersubjectivity, and Imprecision', *The American Journal of Bioethics* 15(6): 16–18.

Savulescu, J. and Kahane, G. (2009). 'The Moral Obligation to Create Children with the Best Chance of the Best Life', *Bioethics* 23(5): 274–90.

Savulescu, J. and Sparrow, R. (2013). 'Making Better Babies: Pros and Cons', *Monash Bioethics Review* 31(1): 36–59.

Singer, P. (1979). 'Not for Humans Only: the Place of Nonhumans in Environmental Ethics, in K. E. Goodpaster and K. M. Sayre (eds), *Ethics and Problems of the 21st Century*. Notre Dame, IN: University of Notre Dame, pp. 191–206.

Sparrow, R. (2007). 'Procreative Beneficence, Obligation, and Eugenics', *Genomics, Society and Policy* 3(3): 43–59.

Sparrow, R. (2008). 'Is it "Every Man's Right to Have Babies if He Wants Them"? Male Pregnancy and the Limits of Reproductive Liberty', *Kennedy Institute of Ethics Journal* 18(3): 275–99.

Sparrow, R. (2010a). 'Liberalism and Eugenics', *Australasian Journal of Philosophy* 89(3): 499–517.

Sparrow, R. (2010b). 'Better than Men? Sex and the Therapy/Enhancement Distinction', *Kennedy Institute of Ethics Journal* 20(2): 115–44.

Sparrow, R. (2011a). 'A Not-so-New Eugenics: Harris and Savulescu on Human Enhancement', *Hastings Center Report* 41(1): 32–42.

Sparrow, R. (2011b). 'Fear of a Female Planet: How John Harris Came to Endorse Eugenic Social Engineering', *Journal of Medical Ethics* 38: 4–7.

Sparrow, R. (2012a). 'A Child's Right to a Decent Future? Regulating Human Genetic Enhancement in Multicultural Societies', *Asian Bioethics Review* 4(4): 355–73.

Sparrow, R. (2012b). 'Human Enhancement and Sexual Dimorphism', *Bioethics* 26(9): 464–75.

Sparrow, R. (2013). 'In Vitro Eugenics', *Journal of Medical Ethics* 40(11): 725–31.

Sparrow, R. (2014a). 'The Real Force of "Procreative Beneficence"', in Akira Akayabashi (ed.), *The Future of Bioethics: International Dialogues.* Oxford: Oxford University Press, pp. 183–92.

Sparrow, R. (2014b). 'Ethics, Eugenics, and Politics', in Akira Akayabashi (ed.), *The Future of Bioethics: International Dialogues*. Oxford: Oxford University Press, pp. 139–53.

Sparrow, R. (2015). 'Imposing Genetic Diversity', *American Journal of Bioethics* 15(6): 2–10.

Suzuki, D. T. and Knudtson, P. (1989). *Genethics: The Clash between the New Genetics and Human Values*. Cambridge, MA: Harvard University Press.

Tooley, M. (1998). 'The Moral Status of the Cloning of Humans', in J. M. Humber and R. F. Almeder (eds), *Human Cloning.* New York: Humana Press, pp. 65–101.

Urban, M. C. (2015). 'Accelerating Extinction Risk from Climate Change', *Science* 348(6234): 571–3.

Zohar, N. J. (1991). 'Prospects for "Genetic Therapy"—Can a Person Benefit from Being Altered', *Bioethics* 5(4): 275–88.

SECTION II

Advancing the Debate

10

Enhancing Conservatism

Rebecca Roache and Julian Savulescu

10.1 Introduction

Recent years have seen polarized debate about novel biotechnologies, including human enhancement, cloning, stem cell research, and the use of PGD to select embryos for non-disease traits. Transhumanists typically welcome developments that aim to overcome human biological limitations, such as our expected lifespan and ordinary levels of intelligence, as some novel biotechnologies promise to do. Bioliberals do not necessarily view such technologies as good but they tend to reject most of the prominent bioconservative reasons for opposing them. For bioliberals, as for political liberals, good reasons to ban novel technologies must make reference to their harmfulness, or the inherent unfairness of their uses. Bioliberals are, as a result, accepting of new technologies on the condition that they are sufficiently safe, fairly distributed, and so on.

Bioconservatives oppose the use of these technologies. While some of their reasons for concern—such as those relating to safety, fairness, and distributive justice—are acknowledged by transhumanists and bioliberals, their main objections are rarely given much weight by their opponents.

Bioconservatives are often political conservatives, and their arguments reflect mainstream conservative thought, including a greater reliance on appeals to intuition than is typical of liberals.[1] Since bioconservatives oppose certain novel technologies even when concerns about harm and fairness do not apply, their views conflict with those of bioliberals and transhumanists.

Bioconservatives oppose enhancement because they believe the goods that enhancement promises to be less valuable than the goods it would destroy. We test this in-principle opposition by considering a hypothetical type of enhancement designed

[1] Parens (2005) argues that the difference between supporters and opponents of enhancement does not correspond to the difference between political liberals and conservatives. However, there is significant overlap between, on the one hand, bioliberalism and political liberalism; and on the other, bioconservatism (to be defined presently) and political conservatism. This chapter focuses on arguments that exist within these overlaps.

to promote values dear to bioconservatives; values that they see as being undermined by enhancement. We will consider how bioconservatives might argue against even this sort of enhancement—that is, against an intervention that promotes bioconservative values by a means that bioconservatives oppose. In doing so, we hope to identify some of the issues most fundamental to the bioconservative case against enhancement.

10.2 Enhancement Policy and Established Wisdom

Leon Kass and Michael Sandel—perhaps the two most prominent bioconservatives, and for this reason the main focus of this chapter—worry that technology develops so fast that we risk embracing it before considering whether what it offers is valuable. Kass remarks that 'this push toward bio-engineered perfection strikes me as the wave of the future, one that will sneak up on us before we know it and, if we are not careful, sweep us up and tow us under' (Kass 2003, p. 10). Sandel writes that '[w]hen science moves faster than moral understanding, as it does today, men and women struggle to articulate their unease. […] The genomic revolution has induced a kind of moral vertigo' (Sandel 2004, p. 1).

Human enhancement—the use of technology (including tools, techniques, and medicines) to raise human capacities above the normal level[2]—especially worries Kass and Sandel.[3] They see it as an unintended, ill-considered side-effect of therapeutic technology. Kass writes:

> It bears emphasis that these powers have not been developed for the purpose of producing perfect or post-human beings. They have been produced largely for the purpose of preventing and curing disease, and of reversing disabilities. […] Yet the 'dual use' aspects of most of these powers […] means that we must not be lulled to sleep by the fact that the originators of these powers were no friends to the Brave New World. Once here, techniques and powers can produce desires where none existed before, and things often go where no one ever intended. (Kass 2003, p. 11)

Similarly, Sandel discusses potentially powerful biotechnologies where '[i]n each case what began as an attempt to treat a disease or prevent a genetic disorder now beckons as an instrument of improvement and consumer choice' (Sandel 2004, p. 2).

Kass and Sandel are concerned that the value of enhancement may be outweighed by the loss of something of greater value. For Kass, this greater value 'may have something to do with what is natural, or what is humanly dignified, or with the attitude that is properly respectful of what is naturally and dignifiedly human' (Kass 2003, p. 17). For Sandel, enhancement 'represent[s] a kind of hyperagency—a Promethean

[2] Specifying exactly what counts as the 'normal level' in this context is itself a difficult task, and one that we do not tackle here. See Bostrom and Roache 2007.

[3] Other bioconservatives who are opposed to human enhancement for reasons similar to those of Kass and Sandel include Francis Fukuyama (2002) and the members of the President's Council on Bioethics (2003), of which Kass was chair between 2001 and 2005.

aspiration to remake nature, including human nature, to serve our purposes and satisfy our desires'; an attitude that 'misses and may even destroy […] an appreciation of the gifted character of human powers and achievements' (Sandel 2004, p. 5). Since enhancement might undermine something of great value, we should resist the temptation to use it.

Both Kass and Sandel hold that, while enhancement should not be embraced if it is unsafe, the most important objections do not relate to safety. Even so, the bioconservative objection that we should avoid enhancement because it might destroy more value than it creates is unconvincing unless bioconservatives have good reason to believe that important values are more effectively promoted by abstaining from enhancement than by pursuing it. Indeed, some bioliberals have argued that some bioconservative values are better promoted by enhancement than by opposition to it.[4]

Traditional conservative thought offers bioconservatives a response to this attack. The caution advocated by Kass and Sandel reflects a typically conservative view that the way things are (or, perhaps, the way things used to be in some bygone era when everything was better) is not the result of current whims or random circumstances; rather, the current arrangements—and in particular the traditions that conservatives esteem[5]—have been shaped by the accumulated wisdom of our ancestors. Let us call this accumulated wisdom of our ancestors *established wisdom*. Reverence for established wisdom, combined with pessimism about the ability of society to withstand radical change,[6] leads conservatives to oppose attempts to overturn the status quo. Bioconservatives can argue, then, that by resisting enhancement we concur with established wisdom. Since established wisdom plausibly exceeds that of bioliberals, it is more likely that bioconservative values will be preserved by resisting enhancement than by enhancing.

This conservative attitude can be traced back to Edmund Burke, often seen as the founder of modern conservative thought. In an influential 1790 essay, he supports the 1688 Glorious Revolution, an event generally viewed as a defence and reassertion of British tradition in the face of a threat.[7] Burke writes of it:

> The Revolution was made to preserve our *ancient*, indisputable laws and liberties, and that *ancient* constitution of government which is our only security for law and liberty. […] The very idea of the fabrication of a new government is enough to fill us with disgust and horror. We

[4] For example, Bostrom 2005, 2008; Savulescu 2007, 2009; Kahane 2011; Pugh, Kahane, and Savulescu 2013; Schaefer, Kahane, and Savulescu 2014.

[5] For some contemporary accounts of the importance of traditions to conservatives, see Kekes 1998 and Scruton 2001.

[6] Kekes 1998, pp. 41–5, 213–19.

[7] In it, King James II of England was overthrown by a collaboration of English parliamentarians and the Dutch William of Orange. Already uneasy about King James's Catholicism and close ties with (Roman Catholic) France, parliamentarians feared the establishment of a Roman Catholic, Anglo-French monarchy when the birth of the king's son in 1688 displaced his Protestant daughter, Mary (wife of William of Orange), as heir to the throne. Following the revolution, William and Mary took the throne as William III and Mary II of England.

wished at the period of the Revolution, and do now wish, to derive all we possess as *an inheritance from our forefathers*. Upon that body and stock of inheritance we have taken care not to inoculate any scion alien to the nature of the original plant. All the reformations we have hitherto made have proceeded upon the principle of reverence to antiquity: and I hope, nay I am persuaded, that all those which possibly may be made hereafter, will be carefully formed upon analogical precedent, authority, and example. (Burke 1790)

Kass and Sandel, like Burke, exhibit 'reverence to antiquity'. Kass bemoans the fact that, '[t]oday, one must even apologise for voicing opinions that twenty-five years ago were nearly universally regarded as the core of our culture's wisdom on these matters' (Kass 1997, p. 18). Sandel remarks that '[i]n order to grapple with the ethics of enhancement, we need to confront questions largely lost from view—questions about the moral status of nature, and about the proper stance of human beings toward the given world'—questions from which 'modern philosophers and political theorists tend to shrink' (2004, p. 2).

It is certainly true that many ideological radical revolutions, such as the various communist revolutions of the twentieth century, backfired. Often, such revolutions have been insufficiently situated in the realities of human psychology and capacity for social life. However, bioliberalism about enhancement is unlikely to involve the sort of radical social change that conservatives typically fear. It is more likely to involve incremental changes as technology advances. Yet Kass in particular seems to object even to gradual social change, as the remarks quoted above suggest. We might take him not only to view radical change as undesirable, but to endorse the stronger view that the status quo represents the ideal state, from which even gradual changes are undesirable. This makes him vulnerable to a charge of status quo bias.[8]

Bioliberals can recognize the value of caution about enhancement. One only has to look around to see that current society celebrates various dubious values. Our consumerist culture depletes our planet's resources at unsustainable levels. Western popular culture celebrates unattainable standards of physical beauty. The reality television craze has led to a prurient obsession with unremarkable people. Reminding ourselves that things have not always been this way helps us keep the status quo in perspective. It is prudent to heed the bioconservative warning not to be seduced by what human enhancement offers without first questioning whether the values it promotes are worthwhile.

Some writers have adopted a questioning attitude in response to the growing popularity of cosmetic surgical enhancement (e.g. Blum 2005; Elliott 2008). Further, some have argued that even when enhancement appears to promote laudable values, promoting these values could have devastating effects in practice. Rob Sparrow argues that, since women have a longer life expectancy than men (along with various other welfare-related advantages), those committed to maximizing the welfare of their children should use sex-selection technology to ensure that their children are female. He

[8] For an account of status quo bias, see Bostrom and Ord 2006.

intends this as a *reductio* of the appealing view that technology should be used in medicine primarily to promote welfare (Sparrow 2010).

Bioliberals would no more welcome the results of implementing Sparrow's sex-selection policy than would bioconservatives, as one of us has argued.[9] Bioliberals can acknowledge that the pursuit of enhancement may be guided by many values that are initially appealing but which must be balanced, and that the premature promotion of a narrow range of values may result in disaster, as bioconservatives like Sparrow fear.

10.3 Does Established Wisdom Really Demand that We Oppose Enhancement?

Even conceding that established wisdom likely outweighs that of today's thinkers, an appeal to established wisdom is insufficient to show that what bioconservatives value is better safeguarded by opposing enhancement than by permitting it. Clearly things *do* change and progress *is* made; history is replete with attempts to improve the human condition, and indeed the human. It is far from clear that the desire to enhance is opposed to established wisdom, rather than reflective of it. The desire to enhance traits such as intelligence, strength, and self-confidence arises from a culture whose established wisdom values these traits.

Further, the process of accumulation of wisdom is ongoing. As such, those who thoughtfully wish to permit enhancement may be viewed as contributing to established wisdom just as our ancestors did. While some modern views—including, according to bioconservatives, the desire to enhance—may appear to contrast starkly with what has gone before, the same can be said of many earlier views that are now accepted as part of established wisdom. In fact, established wisdom is hardly a homogenous, harmonious category of beliefs and practices. Our culture is shaped as much by conflict as by agreement; perhaps even more so. In addition, individuals change their beliefs over time—even bioconservatives. Raymond Vande Wiele became co-director of the first IVF clinic in 1983, yet a decade earlier he had objected so vehemently to IVF that he sabotaged a colleague's experimental attempt to use it.[10]

We might view bioconservatives as adopting a virtue ethics approach, whereby their views about enhancement are shaped by the virtues enshrined in established wisdom. A common criticism of virtue ethics is that it provides insufficient guidance about how to act.[11] When the virtues in question are exemplified by a huge group of people, existing at different times and holding diverse and conflicting views, this difficulty is magnified. It is unclear what the relevant virtues are and how they can guide current decision-making.

[9] Kahane and Savulescu 2010; Savulescu 2015.

[10] See Clarke and Roache 2012, p. 37.

[11] For a discussion, see Hursthouse 1999.

Faced with such problems, a common strategy of bioconservatives, and of conservatives more generally, is to appeal to intuition.[12] While appeal to intuition plays a central role in moral evaluation and reasoning, it plays a dominant role in conservative moral reasoning, since conservatives often resist attempts rationally to analyse and evaluate their moral intuitions.[13] A characteristic conservative appeal to intuition enables bioconservatives to insist both that enhancement conflicts with established wisdom and that it is not possible to articulate exactly how and why this is the case.

This strategy will not work, however. First, different people have different intuitions—for example, bioliberals do not share bioconservatives' intuition that enhancement is bad—meaning that in order to decide who has the right intuitions, we must appeal to some standard of evaluation that is independent of individuals' intuitions.[14] Second, as illustrated by the example of Vande Wiele above, people's intuitions change over time. Kass himself recognizes this when he bemoans that '[w]e have in some sense been softened up to the idea' of cloning since it first came to public attention in the 1960s (Kass 1997, p. 17). Given this, it is likely that even people with an intuitive aversion to enhancement would, over time, become more accepting of it if they were repeatedly confronted with it (e.g. by meeting enhanced people and encountering discussion of it in the media).[15] That they are likely to become more intuitively accepting of it even if it is in fact morally objectionable points to the conclusion that, again, its moral acceptability must be assessed with reference to considerations independent of individuals' unanalysed intuitions.

10.4 Factor X Enhancement: Remaking Humans in the Bioconservative Image

We can emphasize the idea that enhancement need not undermine bioconservative values by imagining a type of enhancement that promotes these values. Following Fukuyama, let us use the term 'factor X' to denote that difficult-to-define aspect of humanity that bioconservatives worry will be undermined by enhancement.[16] Factor

[12] We follow Neil Levy in understanding intuitions as 'spontaneous intellectual seemings', a definition compatible with a range of philosophical and psychological uses of the term.

[13] For a discussion of the role played by intuition in bioconservative arguments against human enhancement, see Roache and Clarke 2009.

[14] Such a standard need not be independent of intuitions in general, however. It could involve, for example, applying the Rawlsian method of reflective equilibrium, in which we conduct an ongoing revision and reconciliation of new and existing beliefs, reasoned judgments, intuitions, and observations in order to maintain a maximally coherent set of beliefs. This process might itself be said to involve intuition.

[15] Clarke and Roache have argued that awareness of intuitions' tendency to change over time can enable liberal governments to introduce technologies such as enhancement into society in a way that maximizes acceptance from conservatives who initially oppose them (Clarke and Roache 2012).

[16] Fukuyama himself is unable to define factor X. He claims that 'when we strip all of a person's contingent and accidental characteristics away, there remains some essential human quality underneath that is worthy of a certain minimal level of respect—call it Factor X' (2002, p. 149), but concedes that 'there is no simple answer to the question, What is Factor X?' (p. 171).

X, let us assume, includes wisdom (esteemed by Kass 2003, pp. 17–19), an appreciation of giftedness (Kass 2003, pp. 19–20; Sandel 2004, pp. 5, 9, 10), human dignity (Kass 2003, p. 20, 2008), humility (Kass 2003, pp. 17, 19; Sandel 2004, pp. 5, 6, 9), 'openness to the unbidden' (Sandel 2004, pp. 6, 9)—that is, the capacity of parents to 'appreciate children as gifts or blessings' (Sandel 2004, p. 6)—and whatever else bioconservatives might view as valuable human features endangered by enhancement. Now, imagine that scientists create a technology that promotes factor X in those who use this technology. Factor X-enhanced people are exactly the sort of people that bioconservatives wish we all were.

Since factor X enhancement, by definition, promotes those features of humanity that bionconservatives value, anticipating how bioconservatives might resist it will be a useful exercise in trying to identify exactly what they find objectionable about enhancement. Even bioconservatives find it difficult to identify the root of their objection, as Sandel notes when he comments that '[w]hen science moves faster than moral understanding, as it does today, men and women struggle to articulate their unease' (Sandel 2004, p. 1), and as Kass acknowledges in his remark that '[i]t is difficult to put this disquiet [about enhancement] into words. We are in an area where initial repugnances are hard to translate into sound moral arguments' (Kass 2003, p. 17).[17]

We'll now consider two reasons that bioconservatives might give for resisting factor X enhancement. First, they might think that enhancement—even of values they endorse—is bad because it is dehumanizing. Second, they might worry that factor X enhancement would increase factor X beyond a level that would be desirable.

Kass and Sandel both fear that enhancement might be dehumanizing (Kass 2003, pp. 10, 15–16, 20, 23, 2008; Sandel 2004, p. 3). Dehumanization is the failure to acknowledge another human (or group of humans) as fully human (Haslam 2006), and a resulting failure to recognize their full moral status.

Since the dehumanized are denied their full moral status as humans, it would be worrisome if enhancement were dehumanizing. Bioconservatives, including Kass and Sandel, worry that enhancement could dehumanize the enhanced, and both bioconservatives and bioliberals have speculated that it could dehumanize the unenhanced.[18] For bioliberals, the worry that enhancement could dehumanize the unenhanced applies only to certain kinds of enhancement—for example, enhancement of cognitive capacity could ground a claim to greater moral status by the cognitively enhanced if

[17] It might be remarked that the best way to find out what bioconservatives think about the idea of factor X enhancement is to ask them. In fact, one of us (Roache) did ask Sandel about something very similar several years ago, in conversation: she asked whether he would object in principle to an enhancement that reliably and safely increased the subject's appreciation of giftedness, a quality he esteems. He responded by doubting that this would be possible. He may be right. Even so, it is likely to be enlightening to explore the bioconservative attitude to factor X enhancement on the assumption that it is (or might be) possible.

[18] For an analysis of this worry, see Savulescu 2009 and Douglas 2013. In speculating about the effects of enhancement on moral status, bioliberals tend not to worry that the unenhanced would lose moral status, but that the enhanced would gain it, and that this would make life worse for the unenhanced. This could happen without the unenhanced being dehumanized, but we gloss over this point here.

cognitive capacity is relevant to one's moral status. It is far from clear that factor X enhancement would constitute such an enhancement.[19] Even so, there is arguably a case for delaying the introduction of factor X enhancement until we are satisfied that it would be a genuine enhancement.

Kass and Sandel, on the other hand, do not distinguish between enhancements that might dehumanize the unenhanced and those unlikely to do so; rather, they fear that enhancement in general might dehumanize the unenhanced. They also fear that enhancement in general might dehumanize the enhanced. For Kass, enhancement would dehumanize the enhanced by undermining human dignity. He sees human dignity as that property that not only underlies our moral status but also gives us our place in the hierarchical natural (and supernatural) order below that of gods and above that of non-human animals (Kass 2008). Unsurprisingly, Kass' view is controversial (Macklin 2003; Pinker 2008), but one need not subscribe to his account of dignity in order to take seriously his worries about dehumanization.

We can compare aspects of Kass' concerns about enhancement to concerns that others have expressed about dehumanization. Kass repeatedly objects to a view of human capacities and achievements that abstracts them from their context in human life. For example, he complains that

> Human experience under biological intervention becomes increasingly mediated by unintelligible forces and vehicles, separated from the human significance of the activities so altered. […] The relations between the knowing subject and his activities, and between his activities and their fulfilments and pleasures, are disrupted. (Kass 2003, pp. 22–3)

By contrast, when expressing his view of human flourishing, he does so in terms that place human subjects firmly in the context of familiar human activities:

> Still, if human flourishing means not just the accumulation of external achievements and a full curriculum vitae but a life-long *being-at-work* exercising one's *human* powers *well* and without great impediment, our genuine happiness requires that there be little gap, if any, between the dancer and the dance. (Kass 2003, p. 23)

He views enhancement as objectionable because it encourages us to view human capacities as abstracted from, rather than embedded in, human activities:

> [O]ne major trouble with biotechnical (especially mental) 'improvers' is that they produce changes in us by disrupting the normal character of human being-at-work-in-the-world, what

[19] Factor X enhancement, as we have envisaged it, involves wisdom enhancement. Enhancing wisdom would almost certainly involve cognitive enhancement (e.g. enhancement of intelligence, memory, and reasoning skills), and as such it could entitle factor X enhanced individuals to greater moral status if cognitive capacity is relevant to moral status. However, wisdom enhancement would likely involve much more than cognitive enhancement. Nick Bostrom, in conversation, has defined wisdom as the ability to get the important things right. This suggests that enhancing wisdom should also involve enhancing moral judgement, which plausibly involves an emotional element. As such, while the wisdom-enhanced would have greater cognitive capacities than the unenhanced, there seems no obvious reason to believe that their presence in society would make things worse for the unenhanced; on the contrary, it could improve life for the unenhanced.

Aristotle called *energeia psyches*, activity of soul, which when fine and full constitutes human flourishing. With biotechnical interventions that skip the realm of intelligible meaning, we cannot really own the transformations nor experience them as genuinely ours. (Kass 2003, p. 24)

Similarly,

There is an old expression: to a man armed with a hammer, everything looks like a nail. To a society armed with biotechnology, the activities of human life may come to be seen in purely technical terms, and more amenable to improvement than they really are. (Kass 2008, pp. 302–3)

This view is echoed again in Kass' comments that 'patients should not be reduced to "thing-hood" or treated as mere bodies' (Kass 2008: p. 301), and in his reference to enhancements as 'peculiar treatments of the body or uses of our embodiments' (Kass 2008, p. 313). He repeatedly contrasts this view of humans as objects or mere bodies with his own view of humans as subjects, deeply rooted in ordinary activities such as 'sewing a dress, throwing a pot, building a fire, cooking a meal, dressing a wound, singing a song, or offering a blessing made in gratitude' (Kass 2008, p. 314).

If enhancement would indeed encourage a view of enhanced or enhanceable capacities as abstracted from the individual humans who have those capacities, Kass could be right to worry. Others have remarked that significant contributing factors to the dehumanization of disabled people are the tendencies to focus on disabilities rather than on individuals who have disabilities, and to define disabled people in terms of their disabilities (Bogdan and Taylor 1989; Haslam 2006; UNICEF 2013).

Because it can be harmful to focus on disability rather than disabled individuals, Kass may be right to worry that enhancement could encourage a similarly harmful focus on (enhanced or enhanceable) capacities rather than on (enhanced or enhanceable) individuals. Kass cites no evidence for his fear that enhancement would encourage this view, but suggestive evidence may be found in the literature on dehumanization. Nick Haslam writes:

The concept of dehumanization features prominently in writings on modern medicine, which is said to dehumanize patients with its lack of personal care and emotional support; its reliance on technology; its lack of touch and human warmth; its emphasis on instrumental efficiency and standardization, to the neglect of the patient's individuality; its related neglect of the patient's subjective experience in favor of objective, technologically mediated information; and its emphasis on interventions performed on a passive individual whose agency and autonomy are neglected. (Haslam 2006, p. 253)

Haslam's references here to medicine's 'reliance on technology', its 'emphasis on instrumental efficiency and standardization', and its 'neglect of the patient's subjective experience in favor of objective, technologically mediated information' should give us pause. Since enhancing ourselves would likely involve increasingly relying on technology, emphasizing instrumental efficiency and standardization (e.g. by altering ourselves in ways that aim to achieve some measurable standard of intelligence, strength, and so on), and focusing on objective, technologically mediated information (e.g. as we

compare our existing capacities with those of others and with the enhanced standards that we wish to achieve), we might reasonably worry that embracing enhancement would exacerbate existing dehumanizing factors in medicine.

Of course, bioliberals would argue—and, indeed, have argued[20]—that enhancement would not dehumanize but would instead amplify much of what is good about being human, and that it is not restrictive but liberating. Recall that bioconservatives too, by definition, endorse the values that factor X enhancement would promote. Even so, those who are worried about dehumanization may insist that what is at issue is not whether enhancement would promote worthwhile values but how it might change public perception of the enhanced and/or the unenhanced. If it would lead people, for example, to focus abstractly on human capacities and their usefulness (or otherwise), and to take an impersonal view of other people as things capable of being improved rather than as individuals, then enhancement could lead to dehumanization.

Whether, and to what extent, introducing enhancement into society would lead people to view themselves and others in this undesirable way is ultimately an empirical matter: we cannot tell antecedently whether it will be a big problem, whether its undesirable effects will be outweighed by the desirable effects of enhancement, whether it can be countered by social measures, and so on. At the very least, bioconservatives might argue, the potentially dehumanizing effects of enhancement call for a cautious approach to it.

10.5 That a Value is Worth Promoting Does Not Entail that it Should Be Made Universal

Bioconservatives may argue that their esteem for factor X does not commit them to endorsing the view that it should be indiscriminately increased. They may instead maintain that while current levels of factor X in society are too low, there exists an ideal level which it would be undesirable to exceed. Given this, if bioconservatives have reason to worry that factor X enhancement might increase factor X in society to an undesirable level, then they have reason to resist such enhancement. This is the problem of overshoot.

One might believe it possible to overshoot the optimal level of factor X if one believes the desirability of increasing factor X in a given society to depend on certain facts about that society, including facts about current levels of factor X. It may, for example, be analogous to the trait of aggression. In a society populated mostly with non-aggressive individuals, aggressive individuals have an advantage because their aggressive strategies are an effective means to win resources. However, if the percentage of aggressive individuals rises above a certain level, being aggressive becomes disadvantageous because one's aggressive behaviour is more likely to result in conflict and injury than in winning resources (Maynard Smith and Price 1973).

[20] See, for example, Pearce n.d.; Bostrom and Roache 2007; Savulescu 2007; Bostrom 2008; Pugh, Kahane, and Savulescu 2013.

It is easier to grasp why too much aggression in society is undesirable than it is to grasp why it might be possible to have too much factor X. Even so, it would not be impossible for bioconservatives to defend the view that there is an optimal level of factor X that it would be undesirable to exceed. One way to do this would be to make the point, noted above, that our society (including the aspects of it valued by bioconservatives) has been shaped by conflict as much as by agreement. Indeed, in capitalist democracies, a diversity of views, insofar as they increase competition and choice, is widely viewed as desirable. As such, bioconservatives might oppose too high an uptake of factor X enhancement in case it reduces constructive and useful conflict by making people too similar. However, while adopting the view that conflict can be desirable may be a reason for bioconservatives to oppose factor X enhancement, it would also require them either to endorse (unless there are independent reasons to oppose) a type of enhancement designed to *increase* desirable conflict, or to argue that desirable conflict in society is currently at an optimum level (Kahane and Savulescu 2014). Failure to do the latter convincingly would leave bioconservative opposition to both potentially conflict-reducing factor X enhancement and conflict-increasing enhancement open to a charge of status quo bias (Bostrom and Ord 2006).

To build a more promising case for the claim that too much factor X in society might be undesirable, bioconservatives might find unexpected support in the liberal views of John Stuart Mill. He argued that individuality and originality are important values that ought to be promoted by society. He viewed them as important in part because they are necessary to the sort of thought that leads to improvement in society:

> There is only too great a tendency in the best beliefs and practices to degenerate into the mechanical; and unless there were a succession of persons whose ever-recurring originality prevents the grounds of those beliefs and practices from becoming merely traditional, such dead matter would not resist the smallest shock from anything really alive, and there would be no reason why civilization should not die out, as in the Byzantine Empire. Persons of genius, it is true, are, and are always likely to be, a small minority; but in order to have them, it is necessary to preserve the soil in which they grow. Genius can only breathe freely in an *atmosphere* of freedom. Persons of genius are, *ex vi termini*, more individual than any other people—less capable, consequently, of fitting themselves, without hurtful compression, into any of the small number of moulds which society provides in order to save its members the trouble of forming their own character. (Mill 1860/1909)

Mill also criticizes 'ape-like imitation', subjugation of oneself to custom and fashion, indifference to individuality, and lack of originality.

This argument suggests that enhancing a significant number of people in the same way may be an undesirable curb on originality, even if the values promoted by the enhancement in question are worthwhile. Kass expresses a similar view:

> We are right to worry that the self-selected non-therapeutic uses of the new powers, especially where they become widespread, will be put in the service of the most common human desires, moving us toward still greater homogenization of human society—perhaps raising the floor but greatly lowering the ceiling of human possibility, and reducing the likelihood of genuine

freedom, individuality, and greatness. […] Indeed, such homogenization may be the most important society-wide concern, if we consider the aggregated effects of the likely individual choices for biotechnical 'self-improvement,' each of which might be defended or at least not objected to on a case-by-case basis. (Kass 2003, p. 16)[21]

While homogenization and stifling of originality may be promoted by certain types of enhancement, however, it is not a necessary feature even of universally adopted enhancement. This is because some enhancements might promote originality and individuality. There are a number of ways in which enhancement might conceivably do this, including reducing the extent to which we are influenced by peer pressure and the opinions of others, increasing creativity, and increasing wisdom. Provided that such values are promoted by factor X enhancement, this type of enhancement would not undermine originality and individuality; on the contrary, it would increase them.

Bioconservatives could respond by arguing that promoting originality and individuality might depend not only on biological changes in enhanced people but also on complex social factors. Mill, after all, writes in the passage quoted above that in order to have geniuses in society 'it is necessary to preserve the soil in which they grow', and that they require 'an *atmosphere* of freedom'. Social factors could prevent factor X enhancement from promoting originality and individuality if, for example, enhanced people were prevented from achieving their potential due to the sort of dehumanizing attitudes discussed above. Whether, and under what conditions, factor X enhancement promotes originality and individuality can only be ascertained empirically and, again, it may be impossible safely to run the sort of large-scale empirical study required to assess this. Such a study would need to run for many generations before we could draw any confident conclusions about the relationship between factor X enhancement and originality and individuality: recall that Mill refers to the desirability of 'a succession of persons whose ever-recurring originality prevents the grounds of [our best] beliefs and practices from becoming merely traditional'. If, at the end of such a long-term empirical study of the capacity of factor X to promote originality and individuality, we found that it is not desirable after all, it would likely be too late to reverse its effects, and we would likely be locked in to the technology.

In general, however, arguments that promoting factor X to too high a level might be undesirable do not justify bioconservatives' outright rejection of factor X enhancement. Bioconservatives clearly believe that there is currently not enough factor X in society and that it ought to be promoted. This view is expressed in their writings. If there is some threshold level of factor X beyond which promoting it is undesirable, bioconservatives have reason to endorse factor X enhancement up to, but not beyond, that level. Of course, it might be that in practice it would be difficult to ensure that we did not overshoot this level. (It is also likely that, in practice, the number of people who would wish to use factor X enhancement would not exceed this level.) However, such

[21] Frances Kamm speculates that a similar concern motivates Sandel's opposition to enhancement (Kamm 2005, pp. 13–14).

observations are—like those relating to distributive justice and safety—of the sort that bioconservatives typically dismiss as secondary to the issue of whether enhancement is truly desirable. In theory, then, appeal to the worry that it is possible to have too much factor X in society will support the bioconservative claim that factor X enhancement should not be adopted beyond a certain level. If they wish to make the stronger claim that it should not be adopted at all, they must appeal to other considerations.

10.6 Conclusion

Often, debate between bioliberals and bioconservatives stalls because the two groups disagree about certain fundamental values, particularly those dear to bioconservatives but less important to bioliberals, such as human dignity and an appreciation of human capacities as gifts. By considering a form of enhancement that promotes exactly what bioconservatives value, and considering whether bioconservatives would have reason to object to such enhancement, we have shifted the focus of the debate away from the question of how valuable the intended results of enhancement are and towards that of whether there is anything wrong with achieving those results through enhancement. We considered two possible grounds on which bioconservatives might object to factor X enhancement. We found that neither supports an outright ban on this type of enhancement. However, the bioconservative concerns that we identified are ones that can be recognized even by those—such as bioliberals—who do not subscribe to typically bioconservative values.

Appeals to originality and individuality to some degree support bioconservative worries. However, to the degree that enhancement promotes such values, it is consistent with the very values that underpin conservative thought. Bioconservatives are surely right that we ought to take serious account of established wisdom and proceed with caution. But we ought not to be paralysed by fear and stuck in the past. Change, including enhancement, can be for the better, in any plausible account of value.

References

Blum, V. L. (2005). *Flesh Wounds: The Culture of Cosmetic Surgery*. London: University of California Press.

Bogdan, R. and Taylor, S. J. (1989). 'Relationships with Severely Disabled People: the Social Construction of Humanness', *Social Problems* 36(2): 135–48.

Bostrom, N. (2005). 'In Defense of Posthuman Dignity', *Bioethics* 19(3): 202–14.

Bostrom, N. (2008). 'Letter from Utopia', *Studies in Ethics, Law, and Technology* 2(1): 1–7.

Bostrom, N. and Ord, T. (2006). 'The Reversal Test: Eliminating Status Quo Bias in Applied Ethics', *Ethics* 116(4): 656–79.

Bostrom, N. and Roache, R. (2007). 'Ethical Issues in Human Enhancement', in J. Ryberg, T. Petersen, and C. Wolf (eds), *New Waves in Applied Ethics*. Basingstoke: Palgrave Macmillan. pp. 120–52.

Burke, E. (1790). *Reflections on the Revolution in France*. London: J. Dodsley, <http://www.fordham.edu/halsall/mod/1791burke.asp>.

Clarke, S. and Roache, R. (2012). 'Introducing Transformative Technologies into Democratic Societies', *Philosophy & Technology* 25(1): 27–45.

Douglas, T. (2013). 'Human Enhancement and Supra-personal Moral Status', *Philosophical Studies* 162(3): 473–97.

Elliott, A. (2008). *Making the Cut: How Cosmetic Surgery is Transforming Our Lives*. London: Reaktion Books.

Fukuyama, F. (2002). *Our Posthuman Future*, New York: Farrar, Straus and Giroux.

Haslam, N. (2006). 'Dehumanization: An Integrative Review', *Personality and Social Psychology Review* 10(3): 252–64.

Hursthouse, R. (1999). *On Virtue Ethics*. Oxford: Oxford University Press.

Kahane, G. (2011). 'Mastery without Mystery: Why there Is No Promethean Sin in Enhancement', *Journal of Applied Philosophy* 28(4): 355–68.

Kahane, G. and Savulescu, J. (2010). 'The Value of Sex in Procreative Reasons', *American Journal of Bioethics* 10(7): 22–4.

Kahane, G. and Savulescu, J. (2014). 'Normal Variation: Refocusing the Enhancement Debate', *Bioethics* 29(2): 133–43.

Kamm, F. (2005). 'Is there a Problem with Enhancement?', *American Journal of Bioethics* 5(3): 5–14.

Kass, L. (1997). 'The Wisdom of Repugnance', *The New Republic*, 2 June, 17–26.

Kass, L. (2003). 'Ageless Bodies, Happy Souls: Biotechnology and the Pursuit of Perfection', *The New Atlantis* Spring: 8–28, <http://www.thenewatlantis.com/publications/ageless-bodies-happy-souls>.

Kass, L. (2008). 'Defending Human Dignity', in the President's Council on Bioethics (ed.), *Human Dignity and Bioethics*. Washington, DC: United States Government Printing Office, pp. 297–331.

Kekes, J. (1998). *A Case for Conservatism*. Ithaca, NY: Cornell University Press.

Macklin, R. (2003). 'Dignity is a Useless Concept', *British Medical Journal* 327: 1419–20.

Maynard Smith, J. and Price, G. R. (1973). 'The Logic of Animal Conflict', *Nature* 246: 15–18.

Mill, J. S. (1860/1909). *On Liberty*, Harvard Classics Vol. 25. New York: P. F. Collier & Son, <http://www.constitution.org/jsm/liberty.htm>.

Parens, E. (2005). 'Authenticity and Ambivalence: Toward Understanding the Enhancement Debate', *Hastings Center Report* 35(3): 34–41.

Pearce, D. (n.d.). 'Brave New World? A Defence of Paradise Engineering'. Available at: <http://www.hedweb.com/huxley/bnw.htm>.

Pinker, S. (2008). 'The Stupidity of Dignity', *New Republic*, 28 May.

President's Council on Bioethics (2003). *Beyond Therapy: Biotechnology and the Pursuit of Happiness*. New York: Regan Books.

Pugh, J., Kahane, G., and Savulescu, J. (2013). 'Cohen's Conservatism and Human Wnhancement', *Journal of Ethics* 17(4): 331–54.

Roache, R. and Clarke, S. (2009). 'Bioconservatism, Bioliberalism, and the Wisdom of Reflecting on Repugnance', *Monash Bioethics Review* 28(1): 4.1–4.21.

Sandel, M. (2004). 'The Case against Perfection', *Atlantic Monthly* 293(3): 51–62.

Savulescu, J. (2007). 'In Defence of Procreative Beneficence: Response to Parker', *Journal of Medical Ethics* 33: 284–8.

Savulescu, J. (2009). 'Genetic Enhancement', in H. Kuhse and P. Singer (eds), *A Companion to Bioethics* 2nd edn. Oxford: Wiley Blackwell, pp. 216–34.

Savulescu, J. (2015). 'Procreative Beneficence, Diversity, Intersubjectivity and Imprecision', *The American Journal of Bioethics*. 15(6): 16–18.

Schaefer, O., Kahane, G., and Savulescu, J. (2014). 'Autonomy and Enhancement', *Neuroethics*. 7(2): 123–36.

Scruton, R. (2001). *The Meaning of Conservatism*, 3rd edn. Basingstoke: Palgrave.

Sparrow, R. (2010). 'Should Human Beings Have Sex? Sexual Dimorphism and Human Enhancement', *American Journal of Bioethics* 10(7): 3–12.

UNICEF (2013). 'See the Child—before the Disability, UNICEF Says', press release, 30 May, <http://www.unicef.org/media/media_69348.html>.

11

MacIntyre's Paradox

Bernadette Tobin

11.1 Introduction

Nearly 40 years ago, Alasdair MacIntyre (1979, pp. 5–7) was asked whether, if it became possible to design our descendants, we should do so. In reply, he said that, if the question were taken to be a question about what would be desirable traits to engender and construct in a whole society (rather than a question about what we should do as parents in raising our own children), then it inescapably raised the problem of identifying what qualities we count as virtues. He argued that, in constructing a 'new table of the virtues', if we succeeded in designing people who possessed these qualities we would have designed descendants who would be so aghast at what we had done to them that they would in turn be unwilling to design their descendants. We would have successfully created descendants who would themselves, by virtue of the very changes we had made to them, reject the whole project. MacIntyre's conclusion, his 'paradox', is striking. It seems right to say, as he does, that if this were the case, then we should not embark on the enterprise of designing descendants. So it may be worth clarifying the nature of his objection to our making large-scale and radical changes to the way we live our lives, and then assessing the strength of this objection to human 'enhancement'. This is what I aim to do.

MacIntyre explains and briefly defends seven qualities which, if we were designing our descendants, we would want them to have:

- the ability to live with uncertainty
- roots in particularity
- the capacity to engage in non-manipulative relations
- the ability to find a vocation in one's work
- the readiness to accept one's own death
- the capacity to hope
- a willingness to take up arms.

He claims that if, as a result of our efforts, these qualities were to come to be at least scattered throughout the population, we would have constructed a society in which

our descendants would not want—as fellow citizens—those who seek to manipulate and control others by non-rational means. They would reject any project to 'treat society as if it were a substance to be moulded'. In addition, we would have contrived for ourselves descendants who, being unable and unwilling to adopt bureaucratic and manipulative modes of planning, would reject the project of designing *their* descendants.

In what follows I will first question whether it might be preferable to assess human enhancement proposals against more mainstream ethics rather than against a 'new table of virtues'. I will then consider whether the seven qualities MacIntyre proposes are indeed virtues. I will then focus directly on the nature of MacIntyre's objection to radical redesigning, clarifying the status of his objection and assessing its pertinence for today's discussions of various proposals to make dramatic changes to human beings.

11.2 Are Virtues a Reasonable Foundation for Assessing Human Enhancement?

The turn to virtue ethics in contemporary moral philosophy, originating in Elizabeth Anscombe's famous article 'Modern Moral Philosophy' (1958), was motivated partly by dissatisfaction with mainstream ethical theory and partly by a recognition of its own merits. Justin Oakley (1996, pp. 128–52) provides a useful summary of both the (continuing) shortcomings of mainstream ethics and the intrinsic merits of virtue ethics.

Mainstream ethics—which is dominated by versions of utilitarianism (albeit sometimes in the more abstract garb of 'consequentialism')—emphasizes questions about what we ought to do at the expense of more global questions regarding what sort of people we ought to be and what sort of lives we ought to lead. It is excessively abstract, addressing itself too little to concrete circumstances and employing an impoverished vocabulary which says little more than 'right', 'wrong', 'obligatory', or 'permissible'. And because of its characteristic emphasis on impartiality, it tends to devalue the ethical significance of particular relationships and to endorse an undesirably minimalist conception of the good life.

By contrast, virtue ethics, particularly of an Aristotelian kind, assumes that the characteristic activity of human beings is the exercise of their rational capacity, and that it is only by living virtuously that the rational capacity to guide our own lives is expressed in an excellent way. In this sense, someone lacking the virtues would not be living a *human* life. The goodness of the virtues is thus based on their connections with essential human characteristics, and they remain good whether or not the agent who has them desires—or would, if suitably informed, desire—to have them. It makes reference to character essential in the justification of an action, and justification according to character takes priority over justification according to rules or results. Thus virtue ethics holds that it is a standard embodied in the idea of human

excellence—rather than a question of whether an action is likely to bring about the best possible consequences—which provides the test of right action. So an account of what makes an action right relies on a prior account of what is valuable in itself, and virtue theory typically recognizes an irreducible plurality in what is humanly valuable or good. Indeed, some intrinsic goods (e.g. an individual's particular friendship with another individual) are agent-relative.

That said, virtue theory need not be thought of as a full-blown alternative to mainstream ethical theory. As Buckle (2002) asks, '[W]hy should being dishonest matter unless dishonesty is wrong?' He thinks that virtue ethics merely emphasizes the fact that *moral* concerns are part of the concerns of the agent: concern for avoiding dishonesty is a support for avoiding lying. In this view, virtue ethics is not so much an alternative to mainstream ethics as a theory of moral right which is supported by a theory of moral psychology and a theory of moral education.

This seems to be how MacIntyre himself thinks of the virtues—that they are qualities which augment rather than replace the other ingredients in the ethical evaluation of acts: prior history, motive, intention, the act itself, the consequences, and the context. He treats the question of the desirability of designing our descendants not as one about the ways in which parents might bring up their children, and not as one about the ways in which we might bring about radical change in each and every person, but rather as one about what would be desirable traits to engender and construct in a society *as a whole*. This recalls Aristotle's view that ethics is the study of the kinds of character trait needed by citizens of a society in which citizen-chosen laws are such as to enable the flourishing of its members individually and collectively. In this view, virtue ethics fills out and augments (and sometimes corrects) mainstream ethics. I shall assume, then, that it provides a reasonable basis from which to assess the wisdom of efforts at human enhancement.

11.3 Are MacIntyre's Seven Qualities Genuinely Virtues?

According to MacIntyre, the question of the desirability of designing our descendants raises a question about what we think would be desirable traits to find scattered throughout a whole society (together with a related question about the ideals or goals these traits are thought to serve). This, he points out, will be, in part at least, a political project. What we will be permitted to do will not be determined solely by the level of our technical ability, nor by the desires of the relevant scientists, nor by those with control over the technologies. It will be determined by what those who hold social and political power decide it is good to do. For this reason the traits we decide to cultivate will determine how far we *ourselves* are able to take and make use of political power. The project of designing our descendants is inescapably also the problem of identifying virtues for ourselves.

Thus MacIntyre's 'new table' of seven virtues. Though one of them is central to his rejection of the project of designing descendants, what he has to say about the others helps to reveal aspects of his overall approach to the question and thus helps to clarify the nature of his objection to making dramatic changes in the way we live. So it will be useful to summarize his account of each quality and to consider whether it recommends itself to us as a virtue, before considering the one he thinks is central to the whole question: a commitment to non-manipulative relations with other people.

The first on his list is a certain kind of flexibility. No trait, says MacIntyre, can be thought of as being good or bad independent of the environment in which it flourishes, towards which it organizes human responses, and from which it derives materials for its tasks. Since the environment, in particular the natural sciences and the technological advances which shape so much of the social environment, is unpredictable, '[w]e will have to design people with all those traits—whatever they may be—necessary for living in an unpredictable environment, people with an ability to live with a large lack of certainty about their future' (MacIntyre 1979, p. 5). It is worth noting that some years earlier, Alvin Toffler's *Future Shock* (1970) marked the birth of the field of 'futurism'. Over 6 million copies of the book were sold and the title became a household term. Though 'futurism' may itself be a dated enquiry, MacIntyre is surely right to say that a certain ability to adapt to technological change is a desirable quality in those who wish to make a contribution to their community. Of course, one does not strictly need that ability—plenty of thoughtful political contributors lack it—but at least an attitude of reflectiveness about the desirability of this or that technological change (e.g. the fact that, because of the phenomenon of 'metadata', perfect strangers, big business, and the government can all connect up with us whether we like it or not) seems to be a desirable quality to have in one's fellow citizens.

However, MacIntyre says, only those who genuinely 'know who they are' will be able to live with this unpredictability 'in a creative way'. Such self-knowledge depends on being able to identify oneself as part of larger projects that contribute to the significance of one's own life, and on being able to identify to whom one belongs, is indebted, owes gratitude, against whom one must stand. Thus the person with MacIntyre's second virtue—'roots in particularity'—is to be contrasted with the universal, cosmopolitan man 'whose moral substance is expressed *only* in an appeal to impersonal, impartial standards in such a way that he is equally at home everywhere in the world so that in fact he is at home nowhere' (MacIntyre 1979, p. 6). Whether one thinks of this quality as a virtue will depend, in part at least, on what one thinks of the claims of social relativism—the view that what counts as a desirable personal quality is tied to the conventions of a particular society. There are more and less persuasive versions of this idea. Certainly cultures have their own ways of instantiating (for example) courtesy towards strangers. But it would be hard to deny that courtesy itself is—universally—a virtue.

MacIntyre's third virtue balances this emphasis on particular allegiances with 'just that kind of impersonality and objectivity that is required if our relationships with

others are to be non-manipulative'. I will defer consideration of this quality—central to his objection to the design project—for the moment.

The fourth trait is that of being able to find some 'productive' work to do in the world. By 'productive', MacIntyre means 'work which has point and purpose both because of the way it involves us in common enterprises with others and because of the ways in which it is worth doing in and for itself'. He argues that thinking only, or even primarily, in terms of outcomes and consequences (pay rates, perhaps, or superannuation arrangements) can be destructive of worthwhile work. And he adds: Providing conditions in which work cannot be done *well* is only less wrong than allowing unemployment. It is hard to see how anyone could reject this virtue, except on the grounds that being able to find any work at all, however non-productive in his sense, will represent progress for those unemployed who wish to find work. Indeed, the idea that the capacity to value work which is productive in MacIntyre's sense provides a telling ground of criticism of contemporary global economic arrangements whereby the living conditions enjoyed by people in affluent countries are made possible in large part because of the living conditions endured by people in poor countries who, even when they do have access to work, hardly engage in work which is 'productive' in this sense.

MacIntyre's fifth virtue is the ability to recognize when one's life's work is done, to accept one's finitude, to accept that a time will come when 'it is right for me to die'. He thinks that the attempt to extend life indefinitely is to intrude on the next generation, to usurp the place of others in the world. Recognizing one's particularity and finitude gives one's life significance and, he thinks, saves one from a destructive desire to remain young. One can agree with MacIntyre on the value of an acceptance of one's finitude at the same time as noting that the desire for longevity does not necessarily involve insensitivity towards the risk of intruding on the next generation. Not all old people are grumpy.

His sixth virtue—the kind of hope found in the ability to trust in goodness even in the face of overwhelming evil—challenges, he thinks, an overly optimistic conception of reason. It involves scepticism towards an Enlightenment-style confidence that the growth of scientific knowledge in and of itself can dispel ignorance. He thinks that the possession of this quality presupposes belief in a reality that transcends what is available as evidence, a belief that rejects merely secular conceptions of reason. But it is possible to conceive of secular counterparts of this attitude. Such an attitude was, I submit, shown by Australians on the occasion of the (then) prime minister's Apology to the Stolen Generation, the generation of children who, for reasons to do with what was then thought to be in their interests, had been removed from their indigenous families. A widespread (secular) response to the apology seemed like an expression of a desire that relations would improve between all Australians, indigenous and non-indigenous: it was a hope-filled commitment.

MacIntyre holds that there is, and has been, no society with worthwhile ends that was brought into being or has preserved itself without war. So his seventh virtue is the kind of courage involved in the willingness to go to war, and to acquire and use the skill

necessary to win a war. This is controversial. Those of us whose lives have been lived entirely in conditions of peace may be hesitant to endorse this claim. Indeed, it is easier to endorse a weaker version of the idea that, if MacIntyre is right about the desirability of all these seven qualities, then *some* kind of readiness to defend a society marked by their presence—against efforts to reshape the society in ways incompatible with them—will be necessary.

The claim that these qualities—together with a commitment to non-manipulative relations with others—are virtues forms the foundation of MacIntyre's rejection of the project of human enhancement. There is plenty of scope for debating the value of this or that quality. True, some of the items on his list are not new. The kind of courage involved in a willingness to defend a society is recognizably Greek, and the ability to trust in the victory of good over evil, even in the face of overwhelming evil, is recognizably both Jewish and Christian. In addition, differences as to what qualities count as human excellences always reflect prior differences about human nature. Think of Plato and Aristotle with respect to temperance (a virtue which seems to be presupposed in MacIntyre's list). They agree it is a virtue but differ with respect to how it is acquired and maintained, and these differences reflect their differing views about human nature (or moral psychology). Indeed, MacIntyre's rejection of the universality and impartiality of contemporary cosmopolitanism, together with his insistence that 'roots in particularity' is itself a virtue, reflects his own view that these features of mainstream ethics presuppose a thinned-out theory of human nature.

But to return to the one quality so far set aside, MacIntyre claims that one of the items in his list will be recognized and valued no matter what one's philosophical theory. He describes it as the kind of moral impersonality that is essential for the creation of a social world in which planning that is both cooperative and rational goes on, a 'kind of impersonality and objectivity that is required if our relationships with others are to be non-manipulative', which is a 'prerequisite if we are to enter relations with others that manifest integrity'. If MacIntyre is right about this, then we can set aside the question of how to judge between the apparently competing claims of very different catalogues of virtues which is raised by different theories of human nature and indeed by differences between the great religions and cultures.

In this regard, MacIntyre says:

> I can urge a particular course of action on you either by seeking to make you an instrument of my own preferences and desires or by appealing to standards (say, of justice or courage or honesty or duty) by which I too am to be judged. Such moral impersonality is essential for two reasons: for the creation of a social world in which the planning is both cooperative and rational, and as a prerequisite if we are to enter into relations with others that manifest integrity. (MacIntyre 1979, p. 6)

Is MacIntyre right when he says that the kind of 'impersonality and objectivity that is required if our relationships with others are to be non-manipulative' is recognized and valued no matter one's philosophical theory? At least since Kant, it would seem so.

Since Kant, two ideas—that (*ceteris paribus*) all of us have an equal ability to see for ourselves what morality calls for, and are in principle equally able to move ourselves to act accordingly regardless of threats or rewards—have come to be so widely accepted that most moral philosophy now *starts* by assuming them (Schneewind 1998). In daily life these two ideas give us the working assumption that the people we live with are capable of understanding and acknowledging in practice the reasons for the moral constraints we mutually expect ourselves and others to respect.

In passing, it is worth noting that MacIntyre's endorsement of a Kantian notion of morality *as* autonomy is different from a contemporary view according to which an individual's autonomy is expressed in *whatever* that person desires, wants, or prefers. The origins of this latter view probably lie in existentialism's focus on the centrality of choice in human experience and its related idealizing of human freedom. Indeed, the idea that what counts as good for a particular person is to be found in whatever that person *chooses* seems to be one of existentialism's legacies. At the very least, this view downplays the idea that there is an objective test in the light of which the worth or value of an individual's decisions and choices can be assessed. When Kant talks about autonomy, he is not talking about autonomy in this way. Though his conception of morality as rational self-governance has made a vital contribution to the rise of the Western liberal vision of the proper relations between the individual and society, an 'existentialist' notion of autonomy is only its pale imitation. So when MacIntyre claims that the kind of 'impersonality and objectivity that is required if our relationships with others are to be non-manipulative' is recognized and valued no matter one's philosophical theory, it is important to note that he has in mind the more traditional Kantian view of autonomy.

11.4 Should We Reject the Project of Designing Our Descendants?

MacIntyre recommends that we reject the proposal to make radical changes to the way we live because, when we identify the qualities we would like in our fellow citizens of the future, the following paradox emerges. We will want them to have the 'seven new virtues' but, just because we value these qualities, we should reject the proposal to design them into our descendants. He gives two reasons. First, we should reject any proposal to treat human society as a substance to be remoulded according to some blueprint. Second, were we to succeed in remoulding our descendants so that they had the qualities we want them to have, they would be deeply ungrateful and aghast at what we had done to them, and would be unwilling to do the same to their descendants. What kind of objection to designing our descendants are these? Are they two objections of the same kind? And how convincing is the claim that they should prevent us from even trying to remould our descendants?

In an analysis of what the 'playing God' objection means, in both its theological and non-theological versions, Coady has distinguished three kinds of objection to

proposals to use science and technology to make dramatic changes in the way we live and cope with the 'many ills, boons, misadventures, and adventures of living today' (Coady 2009). A first kind is based on the idea that a change would be straightforwardly immoral; a second is based on the idea that a change would go beyond the limits of our knowledge, power, and/or benevolence; and a third invokes the idea that a change would transcend the legitimate limits of what it means to be human.

The first kind of objection is straightforward. Think of a modern intelligence officer using science or technology to improve current techniques of torture. The second kind calls attention to three features of human experience: (a) our fallibility (think of the unintended consequences of well-motivated but disastrous social experiments: Coady reminds us of just how comprehensive a disaster has been the introduction of cane toads into Australia to control sugar cane pests, but there are plenty of examples of grand projects (e.g. the damming of the Nile) the results of which are sufficiently mixed to warrant our being less optimistic than we often are about our knowledge); (b) our powerlessness (think of the difficulty we have in shaping institutions which effectively embody our democratic aspirations); and (c) the notorious limits on our benevolence (think of how hard charities find it to engage our sympathies for strangers in remote places who suffer from injustice and bad fortune, and any attempt to alter this by, say, genetic means might have unpredictable and undesirable consequences for our benevolence to those closest to us).

The third kind of objection is that to the process of making radical changes—inherently, in and of itself—rather than to its possible consequences. Coady draws on Nussbaum's (1990) Aristotelian idea—that our perspective on morality is conditioned profoundly by our understanding of what it is for beings like us to flourish—to explain it. He says that '[p]roposals to change human beings […] must themselves be based on moral understandings that are prior to such changes' (Coady 2009, p. 178). So, for example, a proposal to 'make us immune to hazards of love' would damage the moral fundamentals in a way in which a proposal to (say) extend the normal lifespan would not.

MacIntyre may have the second kind of objection in mind when he says that we have reason to reject any proposal to treat human society as a substance to be remoulded according to some blueprint. Of course, sometimes radical social change is for the better. Insofar as is possible, healthcare ought to be as egalitarian in its availability as illness is in its distribution, but it took the radical innovation of a 'national health system' by the Atlee government in England after the Second World War to open at least some eyes to that moral fact. That said, our knowledge of the results of various efforts to remould societies according to *totalitarian* blueprints in the twentieth century is such as to remind us that radical changes according to blueprints are, and ought to be, moderated by a humility born of recognition of our fallibility, powerlessness, and limited benevolence.

MacIntyre certainly has the third kind of objection in mind when he says that, were we to succeed in remoulding our descendants, they would be unwilling to do the same

to theirs. For, as he implies, such a remoulding would, paradoxically, fail the test of 'that kind of impersonality and objectivity that is required if our relations with others are to be non-manipulative'. In making our descendants committed to engaging with others in non-manipulative ways, we would have employed the very interpersonal techniques we would have moulded them to reject.

In calling this a paradox, MacIntyre relies on the Aristotelian idea that the processes of moral development must themselves reflect the character—the moral fundamentals—of what is to be shaped. Remember how Aristotle describes the person who is brave: '[W]hoever stands firm against the right things and fears the right things, for the right end, in the right way, at the right time, and is correspondingly confident, is the brave person: for the brave person's actions and feelings accord with what something is worth, and follow what reason prescribes' (Aristotle n.d., 1115b). That is, mature virtue is acquired not solely by conditioning but also by some process which involves *participation* by the learner.

Though the raw ingredients—appetites and feelings—which are to be shaped into mature virtue are originally (in various ways) 'non-rational', they are susceptible to being modified by reason—not, says Aristotle, by theoretical reason ('arguments don't make people good') but by a process which involves much more than talking. The learner starts by doing what is (for example) brave and gets a taste of a new kind of satisfaction, the satisfaction of finding out that they *can* be brave as well as the satisfaction of *being* brave. No doubt mistakes are made along the way, so the learning involves the kind of cognitive fine-tuning that is embodied in the doctrine of the mean. No doubt there is a place for non-rational correction and encouragement. But the critical thing is that practising *generates insights* along the way. In this way, learners make a contribution to the process itself. Learners are a 'part cause' of their own character, 'begetters' of their own actions, though perhaps Aristotle overstates this when he adds 'as they are of their children' (Aristotle n.d., 1113b). Learners become people who act bravely *from the right motive* (1115b). To put it simply, the process of moral development is a process in which the learner *participates*.

Of course, if someone so lacks the raw ingredients required for this kind of process effectively to get going, and if the consequences of that person continuing to have his or her feelings and volitions unchecked by some such educative process are sufficiently serious, for that person or for others, if he or she is more 'maimed for virtue' than are the rest of us, then a process of behaviour modification by programming or conditioning may be justifiable (and in the worst cases—a violent sociopath comes to mind—society may need to be protected from that person).

When MacIntyre says that, were we thus to design our descendants, they would be ungrateful for what we had done, I think he has in mind the Aristotelian idea that the processes of moral development ought to reflect the character of what is being shaped, and what is being shaped is a set of qualities which cannot be acquired without the *participation* of the learner.

11.5 Conclusion

So should we consider designing our descendants by genetic means? It is sometimes said that our moral resources are so limited, and indeed sometimes so completely absent, that we ought to consider doing this. Indeed, transhumanists say that we should think of human nature as no more than a 'work in progress' (Bostrom 2003). But if MacIntyre is right, then in attempting to make changes to future generations we ought in general to limit ourselves to using only those means, because they require the participation of the learner, are truly *educative*.

References

Anscombe, G. E. M. (1958). 'Modern Moral Philosophy', *Philosophy* 33(124) January, 1–19.

Aristotle (n.d.). *Nicomachean Ethics*.

Bostrom, N. (2003). 'Human Genetic Enhancement: a Transhumanist Perspective', *Journal of Value Inquiry* 37(3): 493–506.

Buckle, S. (2002). 'Aristotle's "Republic" or, Why Aristotle's Ethics is Not Virtue Ethics', *Philosophy* 77(302): 565–95.

Coady, C. A. J. (2009). 'Playing God', in J. Savulescu and N. Bostrom (eds), *Human Enhancement*. Oxford: Oxford University Press.

MacIntyre, A. (1979). 'Designing our Descendants: Seven Traits for the Future', *The Hasting Center Report* 9(1) February, 5–7.

Nussbaum, M. C. (1990). 'Transcending Humanity', in *Love's Knowledge: Essays on Philosophy and Literature*. New York: Oxford University Press, pp. 365–91.

Oakley, J. (1996). 'Varieties of Virtue Ethics', *Ratio* 9(2) September, 128–52.

Schneewind, J. B. (1998). *The Invention of Autonomy: a History of Modern Moral Philosophy*. Cambridge: Cambridge University Press.

Toffler, A. (1970). *Future Shock*. New York: Bantam Books.

12

Partiality for Humanity and Enhancement

Jonathan Pugh, Guy Kahane, and Julian Savulescu

12.1 Introduction

A common strategy that bioconservatives employ in their objections to human enhancement technologies is to appeal to the intrinsic value of human nature, and to object that the use of enhancement technologies might lead us to radically alter this nature. For instance, the President's Council on Bioethics (2003, p. 295) claims that a danger with using such technologies is that in using them we will 'achieve superior results only by compromising our humanity', while Fukuyama (2002, p. 172) similarly warns us against disrupting 'the unity or continuity of human nature' with their use.[1] However, these arguments have been met with scepticism. For instance, critics have pointed out that the concept of human nature that these arguments invoke is ambiguous (and perhaps even indeterminate), and that it is far from clear that changing human nature (if such a thing were possible) must necessarily involve a change for the worse (Fenton 2006; Harris 2007, ch. 3; Buchanan 2009, 2011, ch. 4; McConnell 2010; Nielsen 2011; Pugh 2014). More generally, the standard problem faced by those who appeal to the concept of human nature in their objections to enhancement technologies is to explain why the relatively contingent and arbitrary features of the human species, selected by blind evolutionary processes, should be taken to bear intrinsic value.

In this chapter we shall consider a more plausible strategy for justifying bioconservative opposition to human enhancement that draws on considerations of reasonable partiality. It is commonly claimed that a plausible account of morality must allow for the permissibility[2] of prioritizing certain individuals in one's moral deliberations if one bears a certain kind of special relationship to those individuals (Keller 2013). Perhaps

[1] See also Habermas (2003) and Annas (2009).

[2] One might advance the stronger claim that morality might even oblige us to be partial to certain individuals in our moral deliberations.

more controversially, one might go further and claim that we might also permissibly prioritize in our moral deliberations certain groups to which we belong, even if we do not individually know all members of that group. For instance, some philosophers have argued that we may have reasons to be partial towards our co-nationals in this way (Hurka 1997; Miller 2005; Nathanson 2009).

To extend these considerations of partiality to the human enhancement debate, it might be possible to form a partiality-based objection to enhancement technologies by arguing that we should resist radical departures from human nature, not because human nature possesses intrinsic value but because it is *our* nature, and because we can understand ourselves as having reasons of partiality to preserve our own nature. We shall begin by suggesting that a similar idea can be found in the writings of Bernard Williams and Jerry Cohen, before claiming that their arguments cannot ground powerful objections to many forms of enhancement. We shall then go on to explain the nature of reasonable partiality in more detail, before sketching an account of what partiality for humanity might plausibly involve.

12.2 The Human Prejudice, Conservative Valuing, and Human Enhancement

Both Bernard Williams (2009) and G. A. Cohen (2011) have defended the claim that the fact that we are human beings gives rise to certain sorts of moral reason. We shall briefly delineate each of their arguments in this regard before going on to consider the implications of these arguments for the enhancement debate.

Williams claims that the fact that we are human beings gives us reason to favour human beings in our moral deliberation, even over other species that might share comparably valuable capacities. He starts his defence of this 'human prejudice' by observing that it is possible to trace a humanist tradition throughout history that is based on the fundamental assumption that 'in cosmic terms human beings had a definite measure of importance' (Williams 2009, p. 78). However, Williams himself rejects this tradition in view of the fact that it relies on there being a 'cosmic point of view', or, more prosaically, on there being a way in which things can matter absolutely in the universe.[3] While he denies the existence of a cosmic point of view, he goes on to highlight that this does not entail that there is *no* point of view from which human beings are important. He writes:

> There is certainly one point of view from which they are important, namely ours: unsurprisingly so, since the 'we' in question, the 'we' who raise this question and discuss with others who we hope will listen and reply, are indeed human beings. It is just as unsurprising that this 'we' often shows up within the *content* of our values. Whether a creature is a human being or not

[3] For criticism of Williams' claims about cosmic importance, see Kahane 2014.

makes a large difference, a lot of the time, to the ways in which we treat that creature or at least think that we should treat it. (Williams 2009, p. 80)

In this view, human beings are not understood to have importance in some cosmic, absolute sense. Rather, they simply have importance for human beings.

Williams seems to believe that the mere fact that a creature is a human being can operate as a foundational moral reason to privilege them in our moral deliberations, when a foundational moral reason is understood to refer to a reason that neither admits of nor requires further justification; it is simply something that we care about in a fundamental way. According to Williams,

A central idea involved in the supposed human prejudice is that there are certain respects in which creatures are treated in one way rather than another simply because they belong to a certain category, the human species. We do not, at this basic initial level, need to know any more about them. Told that there are human beings trapped in a burning building, on the strength of that fact alone we mobilize as many resources as we can to rescue them. (Williams 2009, pp. 84–5)

According to Williams, the foundational nature of the moral reason we have to save the human beings in this 'burning building case' suggests that the human prejudice is structurally different from morally reprehensible prejudices such as racism and sexism (Williams 2009, pp. 83–4).

We shall consider the strength of this account, and how it might be used in an objection to human enhancement technologies below. Prior to doing so, it should be acknowledged that Williams' claims seem broadly compatible with some of Jerry Cohen's arguments in defence of conservative forms of valuing, as we shall now outline.

In a posthumously published essay, Cohen sought to defend the conservative attitude of having a 'bias in favour of retaining what is of value, even in the face of replacing it by something of greater value' (Cohen 2011, p. 203). In this attitude, the fact that a bearer of value *exists* can be understood to confer value to that entity in abstraction from the value that it otherwise bears. More specifically, Cohen defends two modes of conservative valuing, which he terms 'particular valuing' and 'personal valuing'. In the case of particular valuing, the object of value in question is understood to bear *intrinsic* value. For example, consider a beautiful landmark such as the Great Pyramid of Giza. By 'particular valuing', Cohen suggests that the value of the extant particular intrinsically valuable object is not exhausted by the intrinsic value which it instantiates; we also value the existing *particular* object that instantiates that value itself, over and above the fact that it possess a certain degree of intrinsic value. To illustrate this distinction, suppose that we could build a grand new version of the Great Pyramid that would bear even more intrinsic value than the original, but that we could only do so by destroying the original (because we had to build the new pyramid on exactly the same spot to ensure the greatest possible aesthetic value). In Cohen's view, the particular value of the extant original pyramid might

give us sufficient reason to refrain from creating the new version, even though it would be better.[4]

In contrast, in the case of personal valuing, the object of such valuing might bear either very little intrinsic value or even no intrinsic value at all; rather, the value of the object lies mainly in its specific relation to the evaluator. To illustrate, suppose that a family had an old dining table that was only just still serviceable. Even though the table might lack intrinsic value and the family could easily replace it with a new one, they might choose to retain it because of their specific relationship to it—that is, because of the table's personal value to them. For instance, the table might have been the scene of memorable discussions, or it might have been built by a much-loved, departed relative. Although Cohen does not point this out, we have suggested elsewhere that in the case of personal valuing, it seems plausible to claim that the relationship of the evaluator to the valued item that grounds personal value is often based in part upon a shared personal history with that item (Pugh, Kahane, and Savulescu 2013, pp. 334, 337).

Both particular and personal valuing can undergird the conservative bias that Cohen seeks to defend. However, it should be acknowledged that he concedes that the conservative bias is defeasible. If the value of the new objects that replace the existing value is of a sufficient magnitude, it may still be correct to bring them about on Cohen's view. However, in such cases, while the conservative may celebrate the new value that has been created, it is also fitting that they should lament the value that has been lost in doing so.

Cohen makes some remarks that suggest that his views regarding conservatism are congruous with the sentiments expressed in Williams' human prejudice regarding the value of human beings. For instance, Cohen writes that we have reason to maintain human beings as they currently exist because they are 'creatures that exhibit a certain form of value', but also (and more pertinently for our purposes here) that we have an additional reason to do so, which is 'that they are *us*' (Cohen 2011, p. 209).

Interestingly, Cohen makes these claims after describing an example that suggests that his views can be readily applied to the enhancement debate:

> If I want us to continue as we are, do I want us to retain our negative features? What if a genetic manipulation could, for example, eliminate envy . . . I would not want to eliminate all of our bad features. I conjecture that that is partly because the negative traits are part of the package that makes human beings the particular valuable creatures that we personally cherish, and are therefore worth preserving as part of that package. (Cohen 2011, p. 209)

He then suggests that this example implies that the particular and personal modes of valuing delineated above correspond to a distinction (alluded to above) between

> *the* reason to preserve human beings—that they are creatures that exhibit a certain form of value, and *our* (additional) reason to do so, which is that they are *us*. (Cohen 2011, p. 209)

[4] Cohen often writes in a way that gives the impression that when we value particulars in this way, we are responding to a special kind of value that goes beyond the intrinsic value they possess. But the claim is better understood as a claim about *how* we should properly respond to the intrinsic value possessed by particular things.

The first reason that Cohen highlights above is grounded by an appeal to particular valuing. Cohen does not explain the content of the 'certain form of value' that he understands human beings to bear (as the above quote intimates), and there are different ways in which this could be cashed out. For instance, it could be argued that human beings bear intrinsic value by virtue of our shared human characteristics or nature. Whichever way we interpret Cohen's claim here, though, the argument that is implicit in his remarks is that even if we could improve human beings in a manner that would increase their intrinsic value, we have reasons to refrain from doing so because human beings as they currently exist have particular value in abstraction from the intrinsic value that they bear (in the same way that the Great Pyramid does). We call this the argument from particular value (Pugh, Kahane, and Savulescu 2013, p. 340).

The second reason that Cohen highlights particularly echoes Williams' claims regarding the human prejudice. In his concluding remarks, Williams claims:

> Hopes for self-improvement can lie dangerously close to the risk of self-hatred. . . . The self-hatred, in this case, is a hatred of humanity. (Williams 2009, pp. 95–6)

For Williams, then, to seek to improve human beings in ways that amount to eradicating the inherent flaws of humankind is to risk expressing a 'hatred of humanity', and to potentially overlook the foundational reason that we have to privilege our (flawed) selves in our moral deliberations. Although Williams does not directly discuss radical forms of enhancement that aim to eradicate flaws inherent in human nature—moral enhancement being a clear example—it seems eminently plausible to expect that he would regard such a project as falling within the scope of his critical remark. Notice, however, an important difference here between Williams and Cohen. For Cohen, the reason to preserve human beings 'because they are us' is not strictly foundational but rather should be understood to correspond to the act of personal valuing. As we explained above, in Cohen's account, the value involved in personal valuing is grounded by the evaluator's relationship to the evaluator, or their shared personal history.

With this in mind, we have suggested elsewhere that there are two ways in which one might argue in favour of preserving human beings as they are by appealing to Cohen's concept of personal valuing. At the individual level, it seems plausible to claim that we each have a relationship to our own features that may be understood as a basis for each of us placing personal value on ourselves. However, to broaden the scope of the argument beyond the individual to the collective level, one would have to claim that humans *as a species* have developed their shared characteristics over the course of a shared biological history. One might then argue that to seek to improve ourselves by using enhancement technologies would be to fail to recognize the significance of our collective relationship to our own shared biological history, and the personal value that we, as a collective, place on it (Pugh, Kahane, and Savulescu 2013, p. 341).

While objections to human enhancement technologies that appeal to conservative modes of valuing or the human prejudice have some advantages over bioconservative

objections that appeal to the much-contested concept of human nature, they are not without their own limitations. We have argued elsewhere that Cohen's arguments concerning conservatism cannot be used to ground a strong sweeping objection to enhancement (Pugh, Kahane, and Savulescu 2013). One reason for this is that, by Cohen's own lights, the conservative bias is defeasible, and we may have reasons to carry out certain enhancements that outweigh the reasons provided by particular or personal value. Furthermore, many attractive enhancements such as life-extending enhancements can be understood as potentially preserving valuable features of human beings in a manner that is compatible with the conservative bias. Indeed, it seems that similar arguments can plausibly be raised against the scope of an objection based on Williams' defence of the human prejudice. As we claimed in our analysis of the application of Cohen's conservatism to the enhancement debate,

> in seeking to change certain aspects of human nature through the use of enhancement technologies, it seems more accurate to say that we are seeking to improve human nature rather than to simply replace it wholesale; and it is not clear why merely seeking to improve ourselves must indicate a hatred of what we now are. (Pugh, Kahane, and Savulescu 2013, p. 352)

The limited scope of the objections to enhancement that can be grounded by Williams' and Cohen's arguments is not their only flaw. Another problem is that the strength of these objections is dependent on the plausibility of Williams' and Cohen's respective accounts of the human prejudice and conservative modes of valuing. Yet it seems that there are problems with both accounts. First, Cohen himself conceded that his account of conservatism faces important objections (Cohen 2011, pp. 214–21). In particular, it is not clear that his conservatism is not simply a manifestation of status quo bias.[5] Furthermore, elsewhere one of us has challenged Williams' defence of the human prejudice by arguing that, despite his claims to the contrary, Williams fails to establish that the human prejudice is morally distinct from other morally deplorable prejudices, such as sexism and racism (Savulescu 2009, pp. 216–35).[6] The piece also includes a defence of the competing 'personism' view that Williams attacks, according to which 'personhood' rather than membership of the human species is what matters morally. In turn, persons are understood to be beings who deserve particularly strong forms of moral protection by virtue of the fact that they instantiate certain valuable characteristics such as rationality and self-consciousness. Although many human beings are persons, not all of them are, as Peter Singer (1979) has famously argued. We shall elaborate on this in Section 12.3.

Rather than attempt to supplement Williams' or Cohen's accounts in order to defend them from these objections here, we shall in Section 12.3 explore a potential approach to human enhancement that shares in the spirit of Williams' and Cohen's comments but is not vulnerable to the objections that have been raised here against their theories.

[5] See also Bostrom and Ord 2006. However, see Nebel 2015 for an argument that the status quo bias can be rational on conservative views.

[6] See also Singer 2009.

More specifically, we shall sketch out the nature and grounds of reasonable partiality in morality, and suggest that this represents a promising way in which bioconservatives might develop Williams' and Cohen's line of argument.

12.3 Introducing Partiality

A 'reason of partiality' is a distinctive kind of moral reason that has its source in a non-instrumentally valued attachment; in the absence of the attachment in question, these reasons would not obtain (Miller 2005, p. 66; Keller 2013, p. 3). Common-sense morality suggests that it can be permissible for agents to incorporate reasons of partiality into their moral deliberations. To illustrate, suppose you are told that you have only enough time to save either two people who are drowning on the left side of a pier, or one person drowning on the right. If you were told that the one person drowning on the right is your brother, this fact might plausibly be understood to make it morally permissible for you to save him rather than the two strangers, even though we would normally claim that you have an obligation to save two people rather than one in such a situation, *ceteris paribus*.

In view of the above, we might claim that we can have (defeasible) reasons of partiality to favour our friends and family in our moral deliberations. However, while this seems intuitively plausible, it raises the question of why certain relationships, projects, and group memberships (which we shall henceforth refer to collectively as 'attachments') can plausibly generate reasons of partiality while others cannot. In the case of family and close friends, this question does not seem to be overly problematic: the close ties that we have forged with these individuals may easily be understood to give rise to a non-instrumentally valuable relationship, and this sort of value can ground reasonable partiality towards our friends and family in our moral deliberations. However, the question is more problematic when we consider partiality towards members of groups to which we belong, because we often lack close ties to all members of these groups. For instance, we surely cannot be morally justified in being partial towards members of certain groups, such as members of our own sex or race, in our moral deliberations. To illustrate, it would be no moral justification of your saving the one instead of the two in the above pier example if you were to point out that the person on the right was the same sex or race as you while the two on the left were not. Call this the 'discriminatory partiality' objection to reasonable partiality towards groups.[7]

To avoid the discriminatory partiality objection, it seems that one must claim that partiality is only morally justifiable towards attachments that are in some way valuable. For instance, John Cottingham (1986, pp. 370–1) suggests that morally justifiable partialities are those in which a plausible case can be made for claiming that the partiality in question must find a place in all or most plausible blueprints for human flourishing. He goes on to claim that partiality to one's own race or sex is not justified in his view

[7] Caney (1996) calls this sort of objection the 'obnoxious identity' problem.

since the available empirical evidence suggests that abandoning partiality to one's race or sex allows members of society greater opportunity to flourish (Cottingham 1986, p. 371).

While Cottingham's account suggests a way in which an account of partiality might avoid the discriminatory partiality problem, the way in which it does so relies heavily on the contingent empirical claim that abandoning partiality to one's race or sex allows members of society greater opportunity to flourish. However, suppose that conditions were otherwise. Cottingham's account seems to suggest that racial and sexual discrimination would be morally acceptable in circumstances in which it did allow members of society greater opportunities to flourish. This seems to weaken this form of response to the discriminatory partiality problem.

To strengthen this sort of account, one would need to establish that there is a non-contingent relationship between human flourishing and the abandonment of racist and sexist attitudes. In such a view, holding such attitudes would simply be inimical to human flourishing. Alternatively one could adopt a view with somewhat broader scope by claiming that partialities are justified as long as they are grounded by a relationship that has final value (Kolodny 2003; Scheffler 1997). In such an account, partialities might be justified by valuable relationships that are finally valuable for non-welfare-based reasons.

12.4 Partiality for Humanity and Human Enhancement

With this discussion of the grounds of reasonable partiality in mind, we can begin to provide a sketch of what partiality for humanity might involve and the implications that it might have for the enhancement debate. We shall first suggest an account of partiality for humanity as being grounded by membership-dependent reasons—that is, by reasons that flow from our non-instrumentally valuing our membership of the human species. Whether or not it makes sense to speak of 'reasonable partiality for humanity' in this sense depends on there being a plausible basis for non-instrumentally valuing 'being human'. If not, then partiality for humanity might be discriminatory.

As we highlighted in Section 12.2, Williams suggests that his burning building case speaks in favour of the human prejudice. However, as one of us has argued elsewhere, deeper consideration of this case suggests that it is not sufficient to establish the claim that we value human beings per se in abstraction from the valuable capabilities that they tend to instantiate, and which demarcate personhood from non-personhood. To see why, consider an analogue of the claim Williams makes on the basis of his burning building case:

> Told that there are *permanently unconscious* human beings trapped in a burning building, on the strength of that fact alone we mobilise as many resources as we can to rescue them. (Savulescu 2009, p. 228)

We believe that this claim has far less intuitive plausibility, precisely because the human beings in question are not persons. Similarly, what seems to be doing the moral work in Williams' original burning building case is the fact that the beings in question are persons.[8]

Consider now two further cases. In the first we must compare the reasons we have to save a human non-person with the reasons we have to save a non-human person:

> *Case 1*: A permanently unconscious human being is trapped in one room of a burning building. A benevolent extraterrestrial who exhibits all the capacities associated with personhood is trapped in another. You only have time to save one.

Ceteris paribus, we believe that you have a moral obligation to save the non-human person in Case 1. The fact that one of the endangered individuals is a person and the other is not gives us sufficient reason to save the former at the expense of the other.

Crucially, though, it is not clear that it is best to construe our moral reason here as a reason of partiality. As we suggested above, we can understand a reason of partiality to be a distinctive kind of reason that has its source in a non-instrumentally valued attachment. However, when we choose to save the person in Case 1, our reason to do so does not seem to stem (at least not primarily) from the fact that we value our membership of the group that contains all and only persons; rather, in this case, our reason to save the person stems directly from their value. The same, it seems, can be said of Williams' original burning building case.

However, the fact that our moral reasons in these cases are not grounded by reasons of partiality to humanity does not entail that such reasons cannot ever obtain. To see why, consider a case in which we must compare the reason we have to save a human person with the reason we have to save a non-human person.

> *Case 2*: A human person is trapped in one room of a burning building. A benevolent extraterrestrial who exhibits all the capacities associated with personhood is trapped in another. You only have time to save one.

In this case it at least seems plausible to claim that one might permissibly choose to save the human person rather than the non-human person. Assuming this is the case, this example suggests that there may yet be scope for a reasonable partiality for humanity that goes beyond the moral reasons that we have to treat persons in certain ways. Case 1 shows that we have strong moral reasons that are grounded in the value of the capacities that are associated with personhood. However, in Case 2, insofar as we might justifiably choose to save the human person rather than the non-human person, the moral reason to do so is not derived from the same set of considerations. Rather,

[8] Some would hold that even if we shouldn't mobilize as many resources as we can to save a human in such a case, we should still mobilize some resources. But this point just brings out the way Williams' example fails to distinguish these two rather different considerations. And as the examples we discuss below show, it is the moral significance of personhood that seems to vastly outweigh the possible further significance of being human, once personhood is set aside.

our reason to do so seems to derive from extrinsic relational factors, or from what we are calling a reason of partiality for humanity.

Our discussion of Cohen's argument from personal value suggests that one possible argument in favour of the view that our membership of the human species is non-instrumentally valuable could be formed by appealing to the fact that humans as a species have developed their shared characteristics, including those that demarcate them as persons rather than non-persons, over the course of a shared biological history. In justifying our choice in Case 2, we might claim that this relationship between you and the human person is a morally relevant factor which, since all else is, *ex hypothesi*, equal, gives you a sufficient moral reason to save the human person rather than the non-human person. In this account, it is not membership of a species per se that is non-instrumentally valuable; as McMahan (2002) also argues, it is difficult to see how this could be morally significant in any way. Rather, what may be understood to ground the non-instrumental value of our membership of the human species is the contingent fact that we as a species have shared a common cultural and biological history with (and crucially *only* with) other members of our species in developing the valuable capabilities that humans instantiate. Although we share parts of our biological history with other species (e.g. chimpanzees), what makes our shared history with other humans special is that only other humans have shared in the part of our history over the course of which we have developed and exercised the very capacities that separate us from other animals.

If this is coherent then there may be some basis for reasonable partiality to humanity that might provide us with sufficient moral justification to save a human person rather than a human non-person in a situation where we have to choose between saving one or the other. However, this partiality provides only very weak reasons. Indeed, it might be claimed that moral reasons that derive from a consideration of the intrinsic properties of other beings might be understood to take lexical priority over our reasons of partiality that are grounded in the non-instrumental value of our attachments. McMahan seems to endorse this sort of view. Although he is sceptical about whether membership of the human species can be non-instrumentally valuable, he writes that, even if this were the case, it would only justify our doing 'marginally more' for them than non-members, and that the baseline for the moral treatment of other species would not be affected by the fact that we have reasonable partiality to human beings (McMahan 2002, p. 227).

The view of reasonable partiality for humanity that we have outlined here is influenced by the arguments of Williams and Cohen that we explored in Section 12.2. However, we believe that it represents a stronger theoretical basis for an objection to human enhancement than the arguments found in Williams' and Cohen's work. First, reasons of partiality are a familiar element of many plausible moral theories. As such, in appealing to such reasons, we do not need to appeal to a novel and controversial form of valuing as Cohen does. Second, in the account that we have outlined above, reasons of partiality are weaker than the reasons that the human prejudice gives in Williams'

account. While they give us moral reasons to favour human beings in our moral deliberations, these reasons only obtain if we have already met our more stringent moral duties to non-human persons that are generated by the value of personhood.

It seems that a suitably developed account of this sketch of reasonable partiality for humanity could provide a theoretical foundation for an objection to human enhancement technologies that would be superior to many existing bioconservative objections that appeal to the concept of human nature. An objection grounded in partiality for humanity offers bioconservatives a way of providing a determinate account of what we have reason to preserve about human beings, without facing the standard problem of explaining how certain relatively contingent and arbitrary features of the human species, selected as they were by a blind evolutionary process, can generate such reasons. In an objection based on the account of partiality illustrated above, the human features that we have reason to preserve are not understood to be intrinsically valuable in the somewhat implausible manner that other bioconservative objections have claimed. Rather, we are understood to have reason to preserve certain features by virtue of the non-instrumental but extrinsic value that these features bear. The extrinsic value they bear is grounded by the fact that these features are the product of our shared biological and cultural history.

Moreover, unlike other objections to human enhancement that appeal to the intrinsic value of human nature, the objection from reasonable partiality is not committed to the claim that changing human nature must necessarily involve a change for the worse. The thrust of the objection is that even if changing human nature might in some ways be a change for the better, we can nonetheless have reason to refrain from making such changes. In the account of partiality outlined above, the reasons at work here are reasons of partiality for humanity grounded in the non-instrumental value of the characteristics that we as a species have developed over a shared cultural and biological history.

However, even if we assume that reasonable partiality for humanity is a plausible position, there are some remaining concerns about whether it can be developed into a successful anti-enhancement argument. The first problem is that in the account outlined above, reasons of partiality are relatively weak and are lexically inferior to the moral reasons generated by the value of personhood. As such, if it can be argued that certain enhancements are necessary to safeguard the moral value of persons, reasons of partiality to humanity would not be sufficiently strong to rule out the moral permissibility of enhancement. Indeed, one of us has argued elsewhere that as our technological capabilities continue to expand, there is an increasing probability that a small group of individuals could destroy the human race, and that the only way to prevent this from happening may be to use technological means to morally enhance the species (Persson and Savulescu 2012).

A second problem with a partiality-based objection to enhancement technologies is that it is not clear that reasons of partiality to humanity can ground a strong objection to forms of enhancement that seek to preserve what is valuable about human beings. For example, we have already argued elsewhere (when considering the application of

Cohen's conservatism to human enhancement debate) that life-extension technologies don't seem vulnerable to this sort of objection because normal ageing brings about the loss of cognitive, sensory, emotional, and motor capacities that we value. Life-extension technologies thus seem to preserve much of what we value in being human rather than replace it with something else. It should therefore be supported by Cohen's conservative bias, and more generally by partiality to humanity (Pugh, Kahane, and Savulescu 2013, p. 345).

12.5 Conclusion

We have suggested that an appeal to reasonable partiality could ground a limited form of objection to human enhancement technologies. We have provided a brief sketch of what a plausible value-based account of reasonable partiality for humanity might involve, suggesting that reasonable partiality for humanity could be grounded by the non-instrumental extrinsic value of the cultural and biological history that we share with all, and only, other members of our species. We have suggested that an objection to human enhancement based on these accounts of partiality has several advantages over common bioconservative objections to human enhancement that appeal to the value of human nature, and the objections that can be drawn from the arguments of Williams and Cohen. Having said this, since reasons of partiality can ground only weak moral reasons, there are some remaining concerns about whether a view of partiality for humanity can be developed into a successful anti-enhancement argument.

We have provided only a brief sketch of how one might give an account of the non-instrumental value of being a member of the human species by appealing to our shared cultural and biological history. However, such an account requires more defence than we can provide here. Indeed, it is telling that both Cottingham and Scheffler are sceptical about whether we can make sense of reasonable partiality for humanity in their value-based account. This in turn raises the worry about whether partiality for humanity may just be 'speciesism' of the sort that Peter Singer (1979) famously attacks by another name. If an account of partiality for humanity cannot stand up to critical scrutiny, even the weak objections to enhancement that we discussed above may simply amount to conclusions that follow from an irrational and morally despicable bias for humanity that is on a par with sexism and racism. Nonetheless, we tentatively suggest that an account of reasonable partiality for humanity represents the most plausible way of developing a bioconservative objection to enhancement based on the significance of human nature, and that it is the most promising way of developing the sentiments expressed in Cohen's and Williams' works. It thus deserves further discussion.[9]

[9] We develop the sketch provided in this chapter in Pugh, Kahane, and Savulescu, 'Bio-Conservatism, Partiality, and The Human Nature Objection to Enhancement' (Forthcoming).

References

Annas, G. (2009). *American Bioethics: Crossing Human Rights and Health Law Boundaries*. New York and Oxford: Oxford University Press.

Bostrom, N. and Ord, T. (2006). 'The Reversal Test: Eliminating Status Quo Bias in Applied Ethics', *Ethics* 116: 656–79.

Buchanan, A. (2009). 'Human Nature and Enhancement', *Bioethics* 23: 141–50, doi:10.1111/j.1467-8519.2008.00633.x.

Buchanan, A. E. (2011). *Beyond Humanity?: The Ethics of Biomedical Enhancement*. Oxford: Oxford University Press.

Caney, S. (1996). 'Individuals, Nations and Obligations', in S. Caney, D., George, and P. Jones (eds), *National Rights, International Obligations*. Boulder, CO: Westview Press.

Cohen, G. A. (2011). 'Rescuing Conservatism: a Defense of Existing Value', in R. J. Wallace, R. Kumar, and S. Freeman (eds), *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. New York and Oxford: Oxford University Press.

Cottingham, J. (1986). Partiality, Favouritism and Morality, *Philos. Q.* 36: 357–73, doi:10.2307/2220190.

Fenton, E. (2006). 'Liberal Eugenics and Human Nature', *Hastings Center Report* 36(6): 35–42, doi:10.1353/hcr.2006.0093.

Fukuyama, F. (2002). *Our Posthuman Future: Consequences of the Biotechnology Revolution*. London: Profile Books.

Habermas, J. (2003). *The Future of Human Nature*. Cambridge: Polity.

Harris, J. (2007). *Enhancing Evolution : The Ethical Case for Making Better People*. Princeton, NJ and Oxford: Princeton University Press.

Keller, S. (2013). *Partiality*. Princeton, NJ: Princeton University Press.

Kolodny, N. (2003). 'Love as Valuing a Relationship', *The Philosophical Review* 112(2): 135–89.

McConnell, T. (2010). 'Genetic Enhancement, Human Nature, and Rights', *Journal of Medicine and Philosophy* 35(4): 415–28, doi:10.1093/jmp/jhq034.

McMahan, J. (2002). *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press.

Miller, D. (2005). 'Reasonable Partiality Towards Compatriots', *Ethical Theory and Moral Practice* 8: 63–81, doi:10.1007/s10677-005-3296-2.

Nathanson, S. (2009). Patriotism, War, and the Limits of Permissible Partiality, *Journal of Ethics*. 13: 401–22.

Nebel, J. M. (2015). 'Status Quo Bias, Rationality, and Conservatism About Value', *Ethics*. 125: 449–76.

Nielsen, L. W. (2011). 'The Concept of Nature and the Enhancement Technologies Debate', in J. Savulescu, R. ter Meulen, and G. Kahane (eds), *Enhancing Human Capacities*. Oxford: Wiley-Blackwell.

Persson, I. and Savulescu, J. (2012). *Unfit for the Future?: The Need for Moral Enhancement*. Oxford: Oxford University Press.

President's Council on Bioethics (2003). *Beyond Therapy: Biotechnology and the Pursuit of Happiness*. New York: Dana Press.

Pugh, J. (2014). 'Autonomy, Natality and Freedom: a Liberal Re-examination of Habermas in the Enhancement Debate', *Bioethics*, doi:10.1111/bioe.12082.

Pugh, J., Kahane, G., and Savulescu, J. (2013). 'Cohen's Conservatism and Human Enhancement', *Journal of Ethics* 17: 331–54, doi:10.1007/s10892-013-9151-0.

Pugh, J., Kahane, G., and Savulescu, J. (Forthcoming in 2016). 'Bio-Conservatism, Partiality, and The Human Nature Objection to Enhancement' The Monist, 95(4).

Savulescu, J. (2009). 'The Human Prejudice and the Moral Status of Enhanced Beings: What Do We Owe the Gods?', in J. Savulescu and N. Bostrom (eds), *Human Enhancement*. New York and Oxford: Oxford University Press.

Scheffler, S. (1997). 'Relationships and Responsibilities', *Philosophy & Public Affairs* 26: 189–209.

Singer, P. (1979). *Practical Ethics*. Cambridge: Cambridge University Press.

Singer, P. (2009). 'Reply to Bernard Williams', in J. A. Schaler (ed.), *Peter Singer Under Fire: The Moral Iconoclast Faces His Critics*. Chicago: Open Court Publishing, pp. 97–102.

Williams, B. (2009). 'The Human Prejudice', in J. A. Schaler (ed.), *Peter Singer Under Fire: The Moral Iconoclast Faces His Critics*. Chicago: Open Court Publishing, pp. 77–97.

13

Enhancement, Mind-Uploading, and Personal Identity

Nicholas Agar

13.1 Introduction

Human biology is a significant barrier to human enhancement. Our biological bodies age. This may, to some extent, be reversible, but it is unlikely to be preventable. There are limits on the capacities of our biological brains to store and process information. The most fundamental of these obstacles arise in respect of the biochemistry of neurons and synapses. This chapter examines a response to these biological barriers to enhancement that involves transferring human identities out of biological brains and bodies into suitably powerful computational devices. Our identity-relevant features are to be uploaded.[1] Suppose we do succeed in rendering all of our identity-relevant capacities in digital form. We should expect them to follow the same steep path of enhancement currently tracked by computers. Humans have no need to fear future autonomous superintelligent robots if we can become them.

In this chapter I shall understand uploading as essentially involving copying. At some point in the uploading process the totality of our identity-relevant features will be copied from biological bodies and brains into new, more robust and powerful vehicles for our identities. I argue that this procedure must inevitably fail because copying does not satisfy a necessary condition for the transmission of human identities.[2] The replacement of human biology by machines may result in significant enhancement, but it will not result in significant enhancement for us.[3]

The most straightforward way to establish the claim that copying cannot transfer human identities would be to defend a philosophically complete theory of personal

[1] The individual associated most prominently with this line of thinking is Ray Kurzweil. See Kurzweil 2005. For illuminating philosophical discussions of uploading, see Koene 2013; Merkle 2013; Chalmers 2014; Corabi and Schneider 2014; Pigliucci 2014; Walker 2014.

[2] This chapter assumes a realist view about personal identity or survival. There are facts of the matter about whether we survive.

[3] In Agar 2014 I dodged the issue of the metaphysics of personal identity. There I argued that reasonably philosophical doubt about whether uploading is survivable makes it prudentially irrational to upload.

identity and argue that it does not permit survival by this means. In what follows I shall refer to the transfer by copying all or part of someone's identity as survival by copying. The problem with this strategy is that our confidence in that conclusion can be only as strong as our confidence in the view about personal identity that justifies it. Philosophers currently defend a variety of theories about personal identity. The friends of uploading who want to respond to this challenge have only to offer philosophical defence of a theory that supports their preferred view about copying. For example, Ray Kurzweil, one of the leading champions of uploading, advances a theory of personal identity that he calls patternism.[4] The pattern currently instantiated by a human brain and body can be instantiated in a computer. According to Kurzweil, this is all that survival by copying requires. Refuting patternism will not refute the possibility of survival by copying so long as some other view of identity that would permit it remains an open possibility.

This chapter takes a more philosophically powerful route to rejecting survival by copying. Rather than defending a full account of personal identity, I defend a necessary condition for the preservation of personal identity or survival. I claim that any plausible account of our identities satisfies this necessary condition, but copying does not. Uploading may copy your identity-relevant features into a computer but it cannot transfer you into it.

13.2 Copying and Teletransportation

This chapter conforms to the time-honoured tradition of philosophical debates about personal identity by beginning with a thought experiment.

The Betelgeuse thought experiment: It turns out that extraterrestrials have located Earth and become fascinated by its human inhabitants. They resolve to study us. The extraterrestrial anthropologists are significantly more moral than are the alien visitors that feature most prominently in science fiction stories. They undertake no abductions of humans and insert no probes into us. The extraterrestrials subscribe to their own version of the *Star Trek* prime directive—they resolve that their studies must not interfere with us either as individuals or as a species. They have technology that permits them to scan and recreate molecule-for-molecule duplicates of Earth's inhabitants. The long-range scanning instruments they use are very sophisticated. Their operation is not detectable by any human technology. The replicas are created in a hidden research facility on an undetected planet of Betelgeuse and are then subject to various experiments. These are not pleasant for the replicas. The need to conduct experiments on relevantly similar test subjects leads the extraterrestrials to prefer that the additional copies be of humans that they copied earlier. Some Earthlings have been copied many times by extraterrestrial researchers, whose clumsiness means that they have an ongoing need for fresh subjects. No one on Earth is ever aware that any of this is going on.

[4] Kurzweil 2005, p. 383.

I suspect that there is a shared intuition that the inmates of the Betelgeuse facility are not fully or partially identical to any Earth-dwelling human. This intuition finds support in the judgement that the copying seems to leave its human targets entirely unaffected. The act of scanning and producing a copy of you can be fully successful without changing the properties and processes that constitute you at the instant you are copied. Put another way, it is epiphenomenal to the events and processes that constitute you or contribute to your ongoing existence.

You might feel sympathy for your copy—if you were ever to discover its existence. But there is no sense in which you should expect to experience the suffering caused to your copy. To put this assessment in a stylistically awkward but more philosophically precise way, none of your time slices prior to a copying event stands in an identity relation to time slices of any occupant of the extraterrestrial research facility. Some philosophers talk about survival rather than identity.[5] In their preferred terminology, none of the time slices of any occupant of the research facility is a survivor of any Earthling prior to the copying event. Both of these verdicts seem compatible with how the extraterrestrials understand the process. The notion that the copies are identical with or survivors of people on Earth would be bad news for the their aspiration to comply with the prime directive. It turns out that their experiments do affect Earthlings in spite of their precautions.

Another thing to notice about the Betelgeuse thought experiment is that it is a variant of the famous teletransporter thought experiment. Teletransportation, a prominent feature of various television series, including *Star Trek*, involves the copying of information about our brains and bodies and its transmission to possibly distant locations where copies are created. However, the teletransporter story is typically told in such a way as to prompt an intuitive verdict, the reverse of that elicited by the Betelgeuse thought experiment. Derek Parfit, foremost among philosophical practitioners of teletransportation, supplements his story about copying with two details absent from the Betelgeuse story. These additions seem to have a significant effect on our intuitions. First, the individual to be copied is aware that they will be copied. Second, once the copy is produced, the original body is destroyed. Parfit predicts that he would feel some anxiety upon entering the cubicle in which his brain and body are to be scanned and vapourized. But any doubts are allayed by the experience of undergoing teletransportation. 'I press the button. As predicted I lose and seem at once to regain consciousness, but in a different cubicle. Examining my new body, I find no change at all. Even the cut on my upper lip, from this morning's shave, is still there.'[6] Parfit is surely right that the individual who emerges from the different cubicle has the strong impression that they are the individual who pressed the button to initiate the process.

Parfit offers the teletransporter thought experiment in support of the psychological continuity account of our survival over time. He suggests that survival, or what matters in survival, reduces to continuities of psychological state. One survives when

[5] See Parfit 1984. [6] Parfit 1984, p. 199.

memories, beliefs, desires, and other psychological states are conveyed from an earlier time slice to a later one. Parfit proposes that any causal connection between psychological states can suffice. In normal survival, psychological connections obtain as a result of the normal operation of the brain. In the Betelgeuse story the alien scanning and copying device does that work. The cause differs but the result is the same. It follows that the Betelgeusians are survivors of Earthlings, or acquire what matters in survival from them. A tally of psychological connections seems to reveal that the Betelgeusian has as good a claim on the identity of the individual about to be copied as the individual who remains on Earth.

Our different intuitive verdicts about the Betelgeuse and teletransporter thought experiments seem to stem from two different perspectives on the copying process. The Betelgeuse thought experiment directs its reader's attention to the experiences associated with the body about to be copied. Nothing seems to happen. In the teletransporter thought experiment the original body is destroyed. Our focus defaults to the mind and body that result from copying. We can continue the Betelgeuse thought experiment in a way that makes the Betelgeusian perspective more salient. Suppose that the narrative switches to a recently materialized human individual convinced that he has been spirited away from Earth to a horrendous alien research facility. His confidence that he is identical to the copied individual is strengthened by the familiar location of this morning's shaving cut.

A standard philosophical strategy at this point would be to offer yet another thought experiment involving copying, and seek to clinch the debate by claiming some philosophical precedent for the intuitive verdict, positive or negative, that it elicits. This chapter takes an alternative path. It resolves the debate in favour of the claim that we do not survive copying by defending a philosophical principle. The perspective of the recently materialized Betelgeusian is understandable but mistaken. It is a case in which reliable means of making judgements about who we are produces false judgements. The sincere feelings of the Betelgeusians that they are identical to residents of Earth is no more veridical than is the sincere belief of a present-day resident of a psychiatric facility that he is identical to Napoleon Bonaparte.

13.3 A Necessary Condition for Survival

I propose that the Betelgeusians are not survivors of any Earthlings. Nor do they acquire what matters in survival from Earthlings. I support this conclusion by first taking a step back to consider some desiderata for a philosophically plausible account of our identities.

Consider the brief period (say five minutes) that commenced with your beginning to read this chapter and that terminates when you have reached this point in the text. Things happened to you. Some brain cells died. A recently consumed meal underwent further digestion. Perhaps some beliefs about identity were altered, though that may be wishful thinking on my part. For most readers of this chapter, nothing significant happened.

By 'nothing significant' I mean nothing that would alter a specific and unique identification of you with a particular body and collection of psychological states. There has been no change to what we might call the 'normal conditions of survival over time'. A change to this specific and unique identification of you with a particular body and collection of psychological states is possible. But something beyond the quotidian events of the type listed in the previous paragraph must have happened for this to be the case.

These normal conditions of our survival over time serve as key reference points for theories of personal identity. Suppose you succeed in formulating a novel theory of personal identity. It had better be the case that your theory endorses survival over the apparently uneventful five-minute time period. It should be considered a major blow against a novel theory that it denies the commonsensical suggestion that our identities are specifically and uniquely connected to particular minds and bodies during that interval. Such a conclusion seems very close to rejecting a realist account of identity or survival over time. Any realist account of personal identity ought to be able to justify the claim that we survive when these normal conditions for our existence obtain. It should not say that our identities come to an end. It should not say that our identities shift from one body to another. It should not say that we split in two with part of our identities moving to another body. It should not say that our identities have fused with the identities of other individuals who commenced the time period attached to some other body.

This is not to say that our identities cannot end, or that we cannot split in two, or that we cannot fuse with some other individual. It is to say that if these things do happen then something quite significant must happen to us. There must be what I will call a significant interference in the properties and processes that normally accompany our survival.

13.4 Significant Interferences in the Normal Conditions of Our Survival

There are differing accounts of precisely which facts make it the case that the individual at the beginning of the five-minute interval is identical to the individual at the end of that interval. Some philosophers will point to psychological continuities.[7] There are only the smallest of changes to autobiographical memories over that time. Perhaps a couple of autobiographical memories were forgotten and a couple were added, but the overwhelming majority are preserved. Others will point to physical continuities.[8] The brain and body at the end of the five-minute interval is physically continuous with the brain and body at its opening. Some philosophers propose that we are identical

[7] For defences of psychological continuity theories, see Parfit 1984; Shoemaker 1984.
[8] For defences of physical continuity theories, see Williams 1973; Unger 1990.

with particular human organisms.[9] The organism at the opening of the interval is the same as the organism at its end. Some philosophers believe in the existence of non-physical Cartesian egos. They will say that the same ego exists throughout.[10]

A complete theory of personal identity requires a sound argument for one of these conditions as necessary and sufficient for survival. But an account of a significant interference requires no such argument. It treats these conditions disjunctively. A significant interference occurs if at least one of these conditions ceases to obtain. It occurs when there is a major disruption in psychological continuity, or when there is a disruption in physical continuity, or when something happens to replace one human organism with a numerically distinct one, or when a Cartesian ego is destroyed or decisively separated from a particular human body.

It is possible to survive significant interferences. Your confidence about whether you survive a given significant interference depends on the extent to which you have rational grounds for believing that what is interfered with is essential to you. Suppose you suffer a mishap that erases all of your distinctive psychology. It deletes all of your autobiographical memories. It eliminates the psychological bases of all of your distinctive values. Otherwise it leaves you unaffected. Does this mishap kill you? Parfit will say: 'Yes'. Advocates of other accounts of identity say: 'No'. According to the view that you are essentially a human organism, it is possible to survive the significant interference of the erasure of your distinctive psychology. Suppose you are essentially a Cartesian ego. You may survive if the post-erasure ego is numerically identical to the pre-erasure ego even if it has shed its distinctive psychology. You should be confident you will survive to the degree that you are confident in the truth of a view about your identity that permits erasure. As your confidence in these views declines, so should your confidence about surviving the mishap. But you should, at a minimum, classify this event as a significant interference in the normal conditions of your survival over time. The mishap is certainly a significant interference, even if philosophical uncertainty about personal identity leaves unresolved the issue of whether or not you survive it.

In this chapter I treat a significant interference as a necessary condition for a change to the specific and unique identification of someone with a particular mind and body. If there have been no significant interferences over a given period then there is a straightforward identification of you to a particular mind and body over that period. You can be confident that you have survived even if you are unsure of the precise philosophical explanation for your survival.

Consider how the search for a significant interference might lead you to reject a claim that something odd has happened to someone's identity. Suppose you have been participating in a seemingly ordinary conversation—perhaps about the state of the weather. Your interlocutor announces that at some earlier point in that conversation

[9] For the view that we are essentially human organisms, see Olson 1997.
[10] For the dualist view of personal identity, see Swinburne 1984.

her existence came to an end—she died. One way to establish the veracity of this unexpected and awkward claim would be to inquire after the theory of personal identity that supports her claim that she has died and ask her to justify it. But we are in a position to offer a more philosophically powerful rebuttal of her claim. You run through a checklist of all of the properties and processes that plausible accounts of personal identity draw on. Were her autobiographical memories abruptly deleted from her brain? Was her brain destroyed and replaced by a duplicate made out of different matter? Did the functioning of some significant part of her brain cease? Did her Cartesian ego either cease to exist or detach itself from her body?[11] If she answers 'No' to each question then you can reject her claim to have died without first determining the truth of a particular account of personal identity. Your confidence in the falsehood of the claim that she has died depends on an incompatibility between that claim and a disjunction of the various theories about how human existence ends that come from philosophically plausible theories of identity.

There can be cases of fission, which occurs when an individual splits in two, or fusion, which occurs when the identity-relevant features of two individuals are combined. Advocates of differing theories of personal identity may prefer different accounts of these. What one theory calls death, another theory may call fission, and another may label straightforward survival. But it should be agreed that for any of these out-of-the-ordinary events to occur, a significant interference is required. Something must happen to the individual as they are currently constituted to cause them to die, split, or fuse. It may be that part of your identity can detach and travel to Betelgeuse. But if no significant interference happened then we can be certain that no part of you travelled to Betelgeuse. The Betelgeuse thought experiment involves copying. This is a causal process, but that process is one-way. The sensitivity of the extraterrestrial scanners means that there is no significant interference with the human individual who is being copied. Perhaps extraterrestrial scanners have certain quantum effects on the psychological or bodily processes of the copied individual. But these effects are not of the type that should suffice to interfere with our identities.

13.5 Old and New Versions of the Psychological Continuity Theory

There seems to be a problem with the purported philosophical modesty of the necessary condition. The most famous contemporary account of survival over time pointedly fails to satisfy it. Parfit (1984) suggests that what matters in survival can be transmitted by an intervention that makes no causal difference to the properties and

[11] Physicalists can accept the idea of change to a Cartesian ego as a necessary condition. If such things do not exist then it is trivially the case that there is no change in the connection between a Cartesian ego and a particular body.

processes that constitute an individual at the time of the putative transmission. We can, according to him, survive by copying.

Parfit's is the most influential current presentation of the psychological continuity theory. Does my exclusion of Parfit's view constitute a problem for the claimed metaphysical modesty of my account? In what follows I argue that Parfit's view is best seen as an eccentric outlier among accounts of personal identity. Other presentations of the view that we are essentially psychological beings satisfy the necessary condition.

The founding insight of the psychological continuity theory is that we go where our minds go. Someone who rejects Parfit's account of how minds are transmitted can accept this basic insight. For example, John Locke describes a case of mind transfer. He says: 'For should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, as soon as deserted by his own soul, everyone sees he would be the same person with the prince.'[12] When the soul and consciousness of a prince comes to 'enter and inform' the body of a cobbler which has been 'deserted by his own soul', the result is that the prince and cobbler swap bodies. They do not swap minds.

As Locke describes this case, it seems to involve significant interference. This is especially clear if we consider the soul to be some kind of immaterial substance that travels between bodies. The soul of the prince detaches from his body. This is a significant interference in the conditions of the prince's existence. Under normal conditions, souls do not detach themselves from bodies. A further change occurs when the soul of the cobbler evacuates his body to make way for the incoming soul. This does not happen in teletransportation, in which part of someone's identity is supposed to arrive in a distant location without ever having departed. This is a feature of the one-way causal nature of copying. Locke's account is likely to endorse the intuitive reaction to the Betelgeuse thought experiment. He will deny that any part of an Earth-dwelling human travels to Betelgeuse unless there is some way for all or part of their soul to make the voyage. This means that the account given in the Betelgeuse thought experiment must be incomplete. A fuller description will include information about how the scanner separates parts of an Earth-dwelling human's soul and transmits it to Betelgeuse, whereupon it comes to occupy the body of a newly created human. Parfit requires no such additions. According to him, what is explicitly stated in the Betelgeuse thought experiment suffices for what matters in the survival of Earthlings to be transmitted to Betelgeuse.

If we reject the existence of immaterial souls, we are deprived of the most obvious way to describe how a particular consciousness might travel from one body to another. Physicalists must say something else about how such transfers could occur. Some physicalist accounts seem to leave open this possibility. According to Jeff McMahan, we are essentially embodied minds.[13] Our survival as embodied minds depends on the continued existence and functioning of our physical capacity for consciousness. He

[12] Locke 1977, p. 185. [13] McMahan 2002.

can accept the idea that we go where our minds go—the transfer of my consciousness from one body to another would bring with it my identity. But this transfer would require the transfer of some significant part of the physical basis of my consciousness. This version of the psychological view satisfies the necessary condition. I submit that any acceptable version of the psychological continuity theory will satisfy the necessary condition. It's not impossible for our minds to switch bodies, but for this to happen there must be a significant interference with us as we are constituted at the time of the switch.

13.6 Fission and Pseudo-fission

A longstanding identity puzzle concerns the possibility of fission. In cases of fission, your identity splits. You go from being a single individual to being two. Philosophers of personal identity have long struggled with the question of how best to describe fission. It seems wrong to say that it kills you, but where do you end up? We can describe cases of fission in such a way that neither post-fission individual has a superior claim on your identity. There's a problem in saying that you are identical to both as they patently are not identical to each other.

I do not propose a solution to this problem here. I limit myself to saying that the necessary condition distinguishes among cases that tend to be grouped together under the heading 'fission'.

One such case involves the surgical separation of the right and left hemispheres of an individual's brain. Both hemispheres are transferred into other bodies that have had their brains removed and destroyed. The original body is destroyed. This is a significant interference. It constitutes a departure from the straightforward and philosophically unproblematic identification of particular individuals with particular minds and bodies. Satisfying a necessary condition leaves open a variety of possibilities about how it affects identity. For example, the view that we are essentially human organisms directs that we take an interest in what happens to the body from which my brain was extracted. If we are essentially human organisms, the procedure will terminate my identity rather than divide it. The intuition that we survive fission would be false.

The philosophical literature tends to place these cases of fission alongside other purported cases that do not satisfy the necessary condition. I reserve the label 'pseudo-fission' for cases like the Betelgeuse thought experiment. The extraterrestrials' creation of replicas involves no significant interference with Earthlings.

Other versions of the copying story do involve significant interferences. Here we are dealing with a composite procedure, part of which satisfies the necessary condition and part of which does not. We must pay attention to which part affects identity. In some versions of the copying story the device does not one but two things. First, it scans the original brain and body and transmits information, permitting a copy to be made. Second, it destroys the original brain and body. As Parfit (1984) tells this story, this combination cancels our intuitive rejection of the idea that we can survive by

copying. If the original is destroyed, our imaginations tend to default to the viewpoint of the being created in the distant location. We have some sense of what it would be like to be that being, to have the impression of abrupt instantaneous travel to an entirely unfamiliar location.

Vaporization significantly interferes with the identities of human individuals. Copying does not. The combination of these two events affects your identity in ways that are no different from vaporization by itself, or vaporization combined with some other activity that does not significantly interfere with the properties that underpin your identity, such as a haircut. A haircut has causal effects on the properties and processes that constitute you. However, like the subatomic effects of the Betelgeusian scanning device, these are not significant interferences. The combination of procedures kills you but not because of anything lethal about haircuts. Copying plus vaporization does not have the effects claimed by Robert Nozick's closest continuer theory of personal identity.[14] It does not cause our identities to magically default to the individual who counts as our 'closest continuer' any more than would a fatal car accident cause your identity to default to your very similar monozygotic twin.

The acts of copying and vaporizing are only contingently connected. We should view their temporal contiguity as a coincidence. As we have seen, in the original version of the Betelgeuse thought experiment there's nothing about the act of copying that requires, or even raises the probability of, the destruction of the original. Moreover, there's nothing that prevents the copy from being produced many millennia after the death of the original. This second possibility requires only that we find a safe place to store the scanned information.

13.7 Explaining (Away) the Intuition that We Survive Copying

We need to say something about the strong intuition that once vaporization eliminates your original brain and body your identity translates to the mind and body in the Betelgeuse research facility. This relates, in large part, to the way we predict that the copied individual will feel about their identity. Parfit picks up on this when he describes the familiar location of the shaving cut on the upper lip. The individual connected with the body created by the facility that receives the transmitted information has the strong feeling that he is identical to the individual who stepped into the teletransporter.

I think that we can explain these mistaken perceptions of identity much in the way that we explain perceptual illusions. We know of many cases in which mistaken beliefs emerge from the operation of a reliable perceptual faculty. We should view copying as generating an identity illusion. It is predictable that certain optical illusions will fool our visual faculties. It is predictable that copying will produce someone with a

[14] Nozick 1981.

confused sense of their identity. The fact that we can describe a thought experiment in which people would tend to make mistakes about their identities does not show that we are not reliable in such judgements. Suppose that you have a large collection of mutually consistent autobiographical memories. You are free of neurological disorders that would interfere with your brain's capacity to generate accurate autobiographical memories. You are almost certainly identical to the person who originally had those experiences. However, just as it's possible to describe circumstances in which your reliable perceptual faculties are fooled into producing false beliefs, so too is it possible to describe circumstances in which your reliable identity sense yields false beliefs about who you are. The example of copying teaches us that accurate identity perceptions could be phenomenologically identical to illusory identity perceptions.

It's not surprising that we might be subject to identity illusions. David Hume famously observed that there was no sense impression corresponding to the self.[15] Psychologists have demonstrated that humans are susceptible to false memory syndrome. We mistakenly attach our identities to sense impressions of events we have never experienced. If these were commonplace and we became aware that they were commonplace then we would question the reliability of our capacity to make accurate identity judgements. Fortunately they aren't.

It's one thing to say that a reliable identity sense can be fooled. What we need is a way to distinguish cases in which a reliable identity sense generates accurate beliefs from cases in which a reliable identity sense produces false beliefs. Fortunately the necessary condition is available to play that role.

Consider a present-day inhabitant of a psychiatric ward who strongly and sincerely believes that he is identical to Napoleon Bonaparte. The first thing we might do to demonstrate the falsehood of his identity belief is to quiz him about his memories. Does he have autobiographical memories of the established events in Napoleon's life? If he has no memory of the events that Napoleon should remember, then that is surely evidence against the identity claim of the psychiatric patient. But we can suppose that during the time of his psychiatric committal he has made an in-depth study of the emperor's life and he has invested much effort in imagining those events. We can further suppose that his research has focused on Napoleon's own testimony, so it is causally connected to the actual events in the emperor's life. This is not so implausible. Research on false memory syndrome indicates that it is possible for people to acquire phenomenologically rich memories that correspond to no past events in their lives.[16] Suppose the intensity of his study gives him memories of the battle of Austerlitz that are causally connected to the emperor's original memories of the battle—he has studied the emperor's own accounts, and, we can suppose, has memories that exactly match the phenomenology of the emperor's memories. At what point in this psychological investigation do we grant that the patient may not be mad and may in fact be Napoleon? I suspect that our reaction to all this effort would be to deepen our concern for his sanity.

[15] Hume 1958, pp. 251–2. [16] Loftus 1975.

The necessary condition described in this chapter supports the harsh assessment that, the faithfulness of the patient's memories to those of the historical Napoleon notwithstanding, he is simply misrepresenting the facts about his and Napoleon's identities. It's not impossible that someone alive today is either identical to or a survivor of Napoleon. But for this to be the case then the creation of this latter-day Napoleon must have significantly interfered with the emperor as he existed from 15 August 1769 to 5 May 1821. Yet it seems implausible that this could be the case here, given that the mind and body that common sense connects with Napoleon's identity were dead and long decomposed before the mind and body that we identify with the psychiatric patient came into existence. The event of his birth could not have significantly interfered with the widely acknowledged life history of the emperor. Sustaining his view requires either backwards causation or that Napoleon's identity is in fact constituted by a non-natural entity that persisted after the destruction of his brain and body. If we don't believe in any of these possibilities we can reject the identity claim of the psychiatric patient. Phenomenological similarities between the memories of the patient and the emperor, and the fact that both sets of memories may trace back to the same historical events, are beside the point. They are further evidence of psychopathology.

Perhaps we should say that to the extent that Napoleon made the modern world, we are all a bit Napoleon. This may be true when we talk about identity in a pointedly metaphorical sense, but it makes no sense in the metaphysically earnest sense that we might appeal to when we wonder if we will survive a forthcoming surgical procedure. At the end of this procedure, will there be someone who is identical with me? There is no one alive today who inherits any of the late emperor's identity in this metaphysically earnest sense.

13.8 Implications for Enhancement

It's possible that we will soon build machine minds much more powerful than current human minds. The mistake lies in the idea that copying our minds into these devices is a way for us to achieve their powers. They will not become us simply by having our identity-relevant characteristics copied into them. This should lead to some caution about the plan to create superintelligent artificial minds. We will not be able to join the revolution in artificial intelligence by copying our identities into the powerful artificial minds that we will create. Since they won't be us, we had better hope that they will play nice with us.

13.9 Concluding Comments

This chapter proposes that the plan to copy a human identity into a machine must fail. For my identity to be transmitted by this process, it must significantly interfere with the

properties and processes that constitute me at the time of the purported transmission. Copying does not satisfy this necessary condition. We are in a position to know that copying does not transmit our identities even when we grant that there is some uncertainty about the processes that underlie our existences over time.

References

Agar, N. (2014). 'On the Prudential Irrationality of Mind Uploading', in Russell Blackford and Damien Broderick (eds), *Intelligence Unbound: the Future of Uploaded Minds*. New York: Wiley Blackwell, pp. 146–60.

Chalmers, D. (2014). 'Uploading: A Philosophical Analysis', in Russell Blackford and Damien Broderick (eds), *Intelligence Unbound: The Future of Uploaded Minds*. New York: Wiley Blackwell, pp. 102–18.

Corabi, J. and Schneider, S. (2014). 'If You Upload, Will You Survive?' in Russell Blackford and Damien Broderick (eds), *Intelligence Unbound: The Future of Uploaded Minds*. New York: Wiley Blackwell, pp. 131–45.

Hume, D. (1958). *A Treatise of Human Nature*, L. A. Selby-Bigge (ed). Oxford: Oxford University Press.

Koene, R. (2013). 'Uploading to Substrate-Independent Minds', in Max More and Natasha Vita More (eds), *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*. New York: Wiley Blackwell, pp. 146–56.

Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking.

Locke, J. (1977). 'An Essay Concerning Human Understanding: Book II', in John Yolton (ed.), *The Locke Reader: Selections from the Works of John Locke*. Cambridge: Cambridge University Press, pp. 176–85.

Loftus, E. (1975). 'Leading Questions and Eyewitness Report,' *Cognitive Psychology*, 7(4): 560–72.

McMahan, J. (2002). *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press.

Merkle, R. (2013). 'Uploading', in Max More and Natasha Vita More (eds), *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*. New York: Wiley Blackwell, pp. 157–64.

Olson, E. (1997). *The Human Animal: Personal Identity Without Psychology*. Oxford: Oxford University Press.

Nozick, R. (1981). *Philosophical Explanations*. Cambridge MA: Belknap Press.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Pigliucci, M. (2014). 'Mind Uploading: a Philosophical Counter-analysis', in Russell Blackford and Damien Broderick (eds), *Intelligence Unbound: the Future of Uploaded Minds*. New York: Wiley Blackwell, pp. 119–30.

Shoemaker, S. (1984). 'Personal Identity: a Materialist's Account', in Sidney Shoemaker and Richard Swinburne (eds), *Personal Identity*. Oxford: Blackwell, pp. 67–132.

Swinburne, R. (1984). 'Personal Identity: a Dualist Theory', in Sidney Shoemaker and Richard Swinburne (eds), *Personal Identity*. Oxford: Blackwell, pp. 1–66.

Unger, P. (1990). *Identity, Consciousness, and Value*. Oxford: Oxford University Press.

Walker, M. (2014). 'Uploading and Personal Identity', in Russell Blackford and Damien Broderick (eds), *Intelligence Unbound: the Future of Uploaded Minds*. New York: Wiley Blackwell, pp. 161–77.

Williams, B. (ed.) (1973). 'The Self and the Future', in *Problems of the Self*, Cambridge: Cambridge University Press, pp. 46–63.

14

Levelling the Playing Field

On the Alleged Unfairness of the Genetic Lottery

Michael Hauskeller

14.1 Introduction

It has become quite common among those who welcome the prospect of enhancing human nature to argue that human enhancement is not only morally permissible but also morally obligatory: that we have a moral duty to enhance ourselves and others, or at least to provide people with opportunities to enhance themselves (e.g. Bostrom 2005; Savulescu 2005; Harris 2007; Persson and Savulescu 2008). This duty is often believed to follow from the fact that our natural talents and abilities are not evenly distributed—that what we can and cannot do is to a large extent the result of a 'genetic lottery' (also referred to as 'natural lottery').[1] In consequence, people's chances in life are, through no fault of their own, hampered by a 'genetic inequality', which, being entirely undeserved, is clearly unfair, and because it is we are morally obligated to do something about it and to, as best we can, 'level the playing field'. Thus Julian Savulescu, Foddy, and Clayton once referred to those who 'do well at sport as a result of the genetic lottery that happened to deal them a winning hand' as the 'genetic elite', arguing that it is unfair

[1] Clearly the term 'natural lottery', which John Rawls in his *Theory of Justice*, first published in 1971, seems to have introduced into the debate on social and political justice, and what exactly such justice consists in and what it requires, is somewhat broader than the term 'genetic lottery'. The natural lottery encompasses all differences between people that are not due to social circumstances or human action, but have come about through natural processes. Whether those differences have been caused by differences between people's genetic constitution or by other factors that influence our development before birth is irrelevant. In contrast, the term 'genetic lottery' emphasizes the presumed cause of those differences and in doing so already suggests the possibility of a solution to the moral problem whose existence the term is meant to express—namely, the unfairness of there being such differences, or of not doing anything about those differences. The solution that is suggested is, of course, some kind of genetic intervention. In this I will mostly use the term 'genetic lottery' because it better conveys the sense that it is indeed possible not only to counterbalance the effects of that lottery but also to tackle and change the lottery itself so that it is no longer a lottery. However, since I do not here attach any importance to the causes of existing natural differences, genetic or otherwise, I will use the term 'genetic lottery' generally in the sense of 'natural lottery'.

that not everybody has the same natural constitution (2004, p. 667). 'Nature', they announced, 'is not fair' (2004, p. 668), which would be uncontroversial if it only meant, as it usually does when we say something like that, that the categories of fairness or unfairness do not apply to nature (i.e. that nature can neither be fair nor unfair), or that nature is not concerned with fairness (but solely with, say, the survival of the fittest, or whatever we think determines the directions that living nature takes). Rather, it is intended to convey the sense that nature is *positively* unfair. It is not just a descriptive statement that means to explain how nature works (or does not work) but instead a normative statement that is intended to imply the moral unacceptability of the natural condition (or certain features of it). Based on this normative understanding, enhancement is then recommended as a remedy against nature's unfairness. 'By allowing everyone to take performance enhancing drugs, we level the playing field. We remove the effects of genetic inequality. Far from being unfair, allowing performance enhancement promotes equality' (2004, p. 668).[2]

Yet the unequal distribution of abilities is not the only kind of unfairness that can be utilized to support the claim that we have a moral duty to enhance. While the above argument makes use of a purely comparative and thus formal understanding of fairness (cf. Hooker 2005), which merely requires that everybody is treated alike or has the same (opportunities, resources, and goods) as everyone else, the moral duty to enhance can also be inferred from a more substantive account of fairness that emphasizes either a) desert or b) need. So even if everyone had exactly the same opportunities, resources, and goods, and was treated in exactly the same way, their situation and treatment might still be unfair if they did not get what they deserved, did not get what they needed in order to live a decent life, or both. Although it is seldom explicitly formulated as an argument, the idea that we—that is, *all* of us—have been treated unfairly *by nature* nonetheless informs much of the pro-enhancement discourse. That idea is at work whenever it is argued that our lives have been blighted by various (natural and hence entirely undeserved) defects, such as certain cognitive and emotional limitations, and the inevitability of ageing and death. This is seen not only as deeply unsatisfactory but, indeed, as unfair, not because there are others out there who have it better but simply because one shouldn't have to live under such conditions, because we didn't do anything (bad) to deserve it and because, frankly, we deserve better. This way of understanding the world and our place in it, which is rather common specifically among transhumanists, is succinctly expressed in Nick Bostrom's claim (which used to grace his personal website[3] but has now disappeared) that once we have enhanced ourselves to a posthuman level, our lives will be 'as they should have been all along'.

[2] It is, of course, highly doubtful that we would do sport a favour by making athletes more equal. As Lisa S. Parker has pointed out (Parker 2012, p. 128), it is the goal of athletic competition 'to establish the inequality of the competitors'. Moreover, the 'presumption of inequalities is integral to sport's entertainment value'.

[3] <www.nickbostrom.com>.

In what follows I will engage mostly with the argument from formal unfairness—that is, the claim that we have a moral obligation to enhance people who are, with regard to their natural constitution, worse off than others *because their being worse off is unfair*. I will concentrate on this argument because it seems to be more widely accepted and because I find it more interesting. However, much of what I am going to say about it is also applicable to the argument from substantive unfairness. Moreover, the argument from formal unfairness will ultimately prove to rely on our tacit acceptance of the argument from substantive unfairness.

14.2 The Genetic Lottery

Moti Mizrahi (2014) has recently argued that natural inequality presents a new problem of evil for the theist. How can a just god allow this to happen? Answering that question may not appear particularly pressing to many of us, but what should be of interest even for those of us who are not theists is the assumption that Mizrahi makes, and that indeed must be made, to support his claim that there is a problem in the first place. That assumption, which Mizrahi treats as obviously true, is that natural inequality is morally obnoxious, and that the reason for that is its unfairness: 'It is important to note that natural inequality is an evil, not because it leads to pain and suffering, although it might and often does, but because it is *unfair*' (Mizrahi 2014, p. 130). Since the talented deserve their talent as little as the untalented deserve their lack of talent, it is, 'from a moral point of view', 'not fair that one person is taller, healthier, faster, thinner, more intelligent, more beautiful, more agile, and otherwise more naturally endowed than another person'. Note that the intended argument does not seem to be that this situation is unfair because God doesn't do anything about it although he could, but rather that God should do something about it because the undeserved inequality of abilities and other desirable traits is indeed unfair.

The argument from formal unfairness clearly draws on the presumed value of equality and raises the question of to what extent we (or God, but that is a different discussion) are morally obligated to promote it. Natural differences between people obviously exist—that is, differences that are not owed to their own choices and actions but to their genetic and epigenetic endowment.[4] Inequality is thus undeniably a basic fact of life. Yet recognizing this fact does not compel us to hold that all undeserved inequality is inherently unfair and needs to be redressed. This is because despite this view's obvious intuitive appeal it is not entirely clear whether it actually makes sense to declare a

[4] It has been pointed out to me that claims that we need to redress the genetic lottery presuppose genetic determinism. However, I don't think that is necessary. I'm pretty sure I could have trained as hard as I liked but I would still never have become a star athlete. That doesn't mean of course that if I had had the right genes then I would have become one. It just means that not everyone is born with the same abilities and the same potential to develop certain abilities. Not everyone can be turned (trained or educated), or can turn (train or educate) themselves, into anything they want. There are limits to what we can achieve in life that are rooted in our individual nature. And that is all, I think, that we need to assume here.

natural condition, which is not in any way the result of human or any other being's agency, to be 'unfair'. Is it unfair that I cannot bend it like Beckham, or play golf like Tiger Woods? We would not normally think so. I may well find it unfair that those people earn so much more money than I, or that they are famous and I am not, based on my conviction that I deserve fortune and fame just as much as they do. But it would be odd to claim that I deserve their *talent* as much as they do. The reason why I might find it unfair that I am not as rich and famous as they are is that my own abilities, even though they are different from theirs, may conceivably appear to me as worthy of being rewarded and publicly recognized as theirs. In other words, my sense of being treated unfairly rests on my conviction that I do in fact possess certain noteworthy reward-deserving abilities myself. Those abilities then form the normative foundation of my claim that I am being treated unfairly. Yet if I lack certain abilities, then I also lack the abilities that I would need to support my claim that I am just as deserving of those abilities as those who actually have them. The problem is precisely that I do *not* have them, and that there are no abilities that I have *prior* to my existence that could possibly serve as a normative foundation for my claim that I deserve certain abilities that others are born with but I was not. I can certainly begrudge others their talent, but that is not the same as finding it unfair that they have it and I do not. It is perhaps psychologically not impossible to do so, and we might even feel tempted to rationalize our grudge and frustration by calling the unequal distribution of talents 'unfair', but it is definitely a stretch,[5] and we know it.

Leaving aside the question of desert, the difference between an unequal distribution of fame and fortune on the one hand and of ability on the other, and the reason why it makes more sense to us to call the former unfair, is that fame and fortune are ultimately bestowed on certain people by human agents, while ability is not. They only get rich and famous because we appreciate their abilities and are willing to reward them. We pay good money to see them play, and we take a keen interest in their lives. If we didn't, they wouldn't be rich and famous at all. So in a way we are giving them their fame and fortune and at the same time withholding it from others. But their abilities are nothing we have given to them. In fact, they have not been given to them by anyone, and that is why it sounds strange to say that it is unfair that they have it and I do not. It would indeed make much more sense if we thought that there is a God or some other agent in the world who is responsible for the distribution of abilities, who actually made a choice and decided, for apparently no good reason at all (because we didn't exist before that decision was made and therefore couldn't possibly have deserved it), that they should have a particular ability and I should not have it. So by claiming that the unequal distribution of abilities is unfair, we actually seem to

[5] If we agree that the unfairness claim presupposes some kind of desert (so that if I don't deserve X, then it cannot possibly be unfair that I don't get X), and that we cannot deserve anything before we are something, because whatever we deserve we deserve on the basis of what we are or do, then it seems to follow that it cannot be unfair that I don't have the same natural abilities as you. Thus the argument is logically valid. Whether we also consider it sound depends, of course, on whether we find the premises persuasive.

imply that there is such an agency in the world—that is, that we have in fact been treated badly by whomever or whatever is responsible for our lack of talent in making the choice to withhold it from us.

This is why the metaphor of the 'genetic lottery' seems peculiarly inapt to convey the view that natural differences are unfair. This is so because lotteries are *not* usually considered to be unfair at all. We all know that if you win a lottery then this is not because you have done anything to deserve it. In fact, the outcome is completely independent of anything you might have done or not done (except taking part in the lottery), or anything you might be or not be. That is why you can be happy about it but not proud. It is completely arbitrary: you were just lucky. You could just as well have not won. This is considered not unfair but, on the contrary, entirely fair. It would only be considered unfair if the odds were stacked—that is, if your win in the lottery were no longer arbitrary but due to someone doing something that made it more likely that you would win the lottery rather than somebody else. In other words, lotteries are considered fair precisely to the extent that their outcomes are arbitrary (i.e. determined by chance) and unfair to the extent that their outcome is not arbitrary (i.e. not determined by chance). We can then infer that the unequal distribution of abilities is perfectly fair, not despite being determined by chance and thus the result of a 'genetic lottery' but precisely for that reason.

14.3 Natural and Social Starting Places

Now it could, of course, be objected that it is not, and was never meant to be, the unequal distribution of natural assets itself that is inherently unfair. Rather, what is, or would be, unfair is not to redress this unequal distribution once it becomes possible to do so. Thus while it is not unfair that some people are less intelligent than others, even if it does restrict their choices in life, it would be unfair to allow this inequality to persist. That is the position that John Rawls suggested in his *Theory of Justice*: 'The natural distribution is neither just nor unjust; nor is it unjust that persons are born into society at some particular position. These are simply natural facts. What is just and unjust is the way that institutions deal with these facts' (Rawls 1999, p. 87). Yet if we accept that we should try to, as much as possible, guarantee fair equality of opportunity, which requires not merely that all legal obstacles to acquiring certain desirable positions in life be removed (careers open to talents) but also that everyone has the same chance to attain those positions (Rawls 1999, p. 63), then it seems that we should also try to distribute the natural goods (e.g. health and vigour, intelligence and imagination) equally since they undoubtedly influence our chances in life just as much as, if not more than, the social conditions we are born into. 'There is no more reason to permit the distribution of income and wealth to be settled by the distribution of natural assets than by historical and social fortune' (Rawls 1999, p. 64). Since those natural inequalities are undeserved—that is, morally arbitrary—we should, according to Rawls, definitely try to redress them by, for instance, spending more money and effort on the education of

those who are less intelligent (Rawls 1999, p. 86). Only very briefly, however, does he entertain the idea that we might try to change nature itself, and instead of merely compensating for a lack of natural abilities make sure that everyone has the abilities they need to pursue a preferred plan of life. For although he admits that it is indeed 'in the interest of each to have greater natural assets', he is unwilling to infer more from this than that society should 'take steps at least to preserve the general level of natural abilities and to prevent the diffusion of serious defects' (Rawls 1999, p. 92), which is still a long way from demanding that natural abilities be distributed equally and ultimately not consistent with the 'level playing field concept of equal opportunity' (Buchanan et al. 2000, p. 65) that Rawls seems to have endorsed.

This inconsistency was addressed and partly corrected by Allen Buchanan and colleagues in their seminal and aptly named treatment of the issue, *From Chance to Choice*:

> If precise and safe control over the distribution of natural assets becomes feasible, then those who believe that justice is concerned with the effects of natural assets on individuals' life prospects will no longer be able to assume that justice requires only that we compensate for bad luck in the natural lottery by intervening in the social lottery, rather than by attacking natural inequalities directly. (Buchanan et al. 2000, p. 64)

If the goal is to create a situation where everybody starts life with the same opportunities, then it doesn't seem to matter whether what stands in the way of that goal is a certain social structure or an uneven distribution of natural assets. In fact, once we have learned to manipulate our nature, the distinction between social assets and natural assets becomes blurred, if not altogether untenable (Parker 2012, p. 128). The same holds for the distinction between social inequalities and natural inequalities (Lewens 2009). It seems, therefore, that there is a case to be made for the removal of natural inequalities (Holtug 1999, p. 139),[6] and some have indeed made this case very forcefully. Mark Walker, for instance, has argued that since people who are naturally very happy (hyperthymic), because they are genetically predisposed to have a higher average level of happiness than most people, tend to have much better opportunities in life, we have a moral duty to provide those who were less fortunate in the genetic lottery with the means to acquire the same high level of average happiness. That duty is derived from the alleged unfairness of the fact that not everybody is as happy as the hyperthymic:

> To deny the rest of us access to HPP (Happy People Pills) is a grave form of injustice, for it is to artificially limit the pool of this valuable resource; and to restrict the pool of this valuable resource is to prohibit most of us from the opportunity for what many (but not all) of us see as the best life: life with the happiness and achievement of the hyperthymic.' (Walker 2009, p. 35)

[6] Torbjörn Tännsjö (2009, p. 325) has even suggested that the reluctance to support and promote enhancement for those less favoured by nature is owed to a 'Nietzschean view of justice, according to which it is unfair if those who are less fit pool their resources and rob the genuinely strong Übermensch of his genetic advantage'. Needless to say, Tännsjö recommends that we get rid of this antidemocratic and indeed 'fascistoid' notion of justice as quickly as possible.

However, if those who are naturally happy or hyperthymic did nothing to deserve such happiness, then it is hard to see why we should think that those who are not naturally happy deserve to be happy. It is true that we do not deserve to be unhappy, but we do not *deserve* to be happy either. Yet if we neither deserve to be unhappy nor deserve to be happy, then it seems that considerations of fairness simply do not apply. It would only be unfair to leave people in their (after all only comparatively) unhappy constitution if it were unfair for them to be in that condition in the first place. Thus the argument from formal unfairness presupposes some degree of substantive unfairness. There are two ways in which the claim that it is substantively unfair that not everyone is by nature extraordinarily happy could be plausibly defended. It could be unfair if a) we all deserved to be extraordinarily happy or b) we needed to be extraordinarily happy to live a decent life—that is, one from which none of the basic goods are missing. Yet since nobody needs to be hyperthymic to live a good life and there is no good reason to suppose that we deserve to be better than we actually are, the argument from substantive unfairness also appears unconvincing.

14.4 Equality with Whom?

Let us assume for the moment that we accept the above argument in favour of enhancement for the sake of equal opportunity. If, say, A is by nature more intelligent or more happy than B, so that A has a better chance of getting on in life, we should either compensate B for their undeserved handicap in the game of life, or, provided that it is possible, lift B up to the intelligence or happiness level of A so that both have the same starting conditions. Those who show themselves to be persuaded by this kind of argument usually suppose A and B to be humans. But is there any good reason why we shouldn't apply the underlying principle to humans and non-humans alike? Consider the following argument. I cannot fly, but if I could fly I would have many opportunities that are not open to me now. It would also give me great pleasure. Thus the quality of my life would arguably be enhanced if I could fly. Yet there are other living beings that can fly, such as birds. This is a result of the genetic lottery which has given birds abilities that have been denied to me. That is unfair because the fact that birds can fly and I cannot is morally arbitrary. Birds did nothing to deserve their ability to fly, as I have done nothing to deserve my inability to fly. Therefore society has a moral obligation to provide me with the means to acquire that ability, should such means be either already available or at least within our reach. Since there are many different species of animal, most of which have abilities that humans do not currently possess, there is obviously a lot to enhance before we will actually have levelled the playing field completely. Bats can echolocate; dogs have a sense of smell that is 10,000 times as acute as ours; ants can carry loads that exceed their own body weight multiple times; and so on and so forth. Is all this unfair? Are we morally entitled to be able to do what they can do? I suspect that most of those who believe that we have a moral duty to level the playing field by enhancing people who seem to have lost out in the genetic

lottery would not want to go that far. But why should we regard the species boundaries as relevant here? Why should the genetic lottery suddenly cease to be unfair (and thus in need of correction) when we cross the species barrier?

The argument works, of course, in both directions. Not only are there many things that animals can do that we cannot do, but there are also things that we can do that animals cannot do. So should we also level the playing field for them? Nick Bostrom (2004) may have been the first who suggested as much in his whimsical fantasy about Albert, the uploaded and cognitively enhanced golden retriever who is interviewed on *Larry King Live*. After reflecting on the unequal distribution of luck among both humans and animals, the former dog, now post-dog, Albert declares himself to be the 'lucky one', and when asked by Larry King whether he had a mission in life, he replies that he wants 'everyone to be the lucky one'. Given the context, this can only mean that he wishes for all animals to be subjected to the same cognitive enhancement process that he himself has undergone. That animal enhancement or 'uplifting'[7] is indeed a matter of distributive justice has been expressly argued by George Dvorsky (2008). He claims that we have a duty to cognitively enhance animals because not to do so 'would be an unfair distribution of primary goods that are requisites for political participation, liberty and justice'. This is thought to be unfair because it is the result of pure luck: 'Like some humans who argue that they have fared poorly in the genetic lottery, it can be said that nonhumans have missed out in the species lottery' (2008, p. 136). Others, such as Sarah Chan (2009), have also argued that we have a moral obligation to enhance animals, although not directly on the grounds that we would be treating them unfairly if we didn't. Rather, the claim is based on the fact that animals have interests and that they matter morally just as much as ours. However, it seems to me that this argument ultimately also appeals to our sense of fairness: the reason why we should take into account not only human interests but also the interests of animals is, presumably, that unequal treatment is not morally justifiable. In other words, it would be unfair to take heed of human interests but disregard those of animals.

Yet if we do level the playing field so thoroughly that everyone (be they human, animal, or machine) has all the abilities that everyone else has, so we have ended the tyranny of the genetic lottery, then all differences would have disappeared. It would be a thoroughly equalized (*gleichgeschaltet*) world. Why would we want to live in such a world? Differences clearly also have their value. They make life interesting and rich. They give us an individual identity. While a certain degree of equality may be desirable,

[7] The use of the word 'uplifting' in this context is quite telling. It suggests, even more than the word 'enhancement', the possibility of a linear progression through a series of clearly defined stages or steps that reach from the bottom to the top. Each entity is situated on a particular step, and when it is 'uplifted' it progresses to one of the next steps. The higher you are on that stepladder, the better you are. Animals are assumed to be on a lower step than humans. Uplifting them might raise them to the same step on which we currently find ourselves. Further uplifting would raise them to a higher step, which would make them superior to us, unless we also progressed to the next stage.

universal sameness certainly is not. Fortunately it is also highly unlikely, if only because it may be impossible to have *all* the abilities that can currently be found in any individual. Some of those abilities are likely to be mutually exclusive, so that you cannot have the one and the other. Levelling the playing field completely would then be not only undesirable but also impossible.

14.5 Equality of What?

The same holds, of course, if we restrict our ambitions to the human world. I cannot bend it like Beckham or play golf like Tiger Woods, no matter how hard I try. I just don't have the ability. Nor can I play the violin like Yehudi Menuhin or paint like Picasso, no matter how hard I try. This is no fault of my own and therefore morally arbitrary. But it is hard to see how we should be able to have all those different abilities combined in one and the same person. Should we then try to level the playing field to the extent that it is possible?

The problem with this suggestion is that before we can start attacking natural inequalities, we first need to clarify what exactly we intend to make equal. Arguably it is chances in life, but it is not always obvious which natural assets increase our chances in life, and how much of any of those assets we need to have to have better life prospects. One should think that being smart or good-looking is not exactly a disadvantage in the game of life, but there is no evidence that it is the smartest people who are economically most successful,[8] or, even more importantly, who are most likely to have the best or happiest life (Veenhoven and Choi 2012). And how smart exactly do we have to be? It is certainly not the case that the smarter you are the more opportunities in life you will have and the better your life will be. Whether a so-called natural asset is indeed an asset for you depends to a large extent on the circumstances, not only those created by the reigning social structure, but also those that depend on what other abilities you have, and what kind of opportunities emerge in the course of your life. 'Thus, there is no such thing as a resource per se. Different traits will be resources in different social environments. Recognition of this simple fact imposes a fundamental constraint on any attempt to intervene in the natural lottery in the name of equality of opportunity or resources' (Buchanan et al. 2000, p. 80).

In some cases it is even difficult to know whether having an ability or not having it makes for better opportunities in life, and hence whether it is unfair that you have it or unfair that you do not have it. If, for instance, you have few or no moral scruples and are prone to stealing stuff and killing people, you are likely to end up in prison, which is clearly not fair since you have never asked to be bad in the first place. It is just your bad luck of being born with the wrong genes and into the wrong social circumstances.

[8] <http://www.forbes.com/sites/joshlinkner/2013/02/14/6-reasons-the-smartest-people-arent-the-most-successful/>.

Hence, since you haven't really chosen to be bad,[9] we owe it to you to compensate you or, if we have a way to do that, to change your nature in such a way that you are no longer bad. We have a moral obligation to make you good. But then again we could just as well argue the opposite. A lack of moral scruples, or let us say a certain moral flexibility, is not always a disadvantage in life. It is, after all, often people with moral integrity who miss out on opportunities that might make their life better, at least in the sense that they might give them access to certain goods and resources, and increase their liberty. So it seems that we owe it to you to make you less good, so that you are not disadvantaged in the game of life by your moral scruples.

This shows two major problems with the whole argument from formal unfairness. The first is that equality of opportunity may not be the only relevant value at stake here. Perhaps it is better for people to have moral integrity even though it does diminish their ability to acquire certain goods. The other problem is that we don't always know what exactly enables us to have a good life, and that is in fact the fundamental problem with the whole enhancement project (Hauskeller 2013). If it would not necessarily and obviously be better for me to be as intelligent as those who are more intelligent than I, then there is no clear reason to think that the situation is unfair. In order for the fact that you have something that I do not to be considered unfair, what you have must (as a necessary condition) be in some sense better than what I have. It must be considered beneficial or in some other way worth having, but whether or not it is worth having depends on many contingencies as well as on our own individual idea of a good life.

14.6 Is Life a Race?

It is, of course, not entirely impossible to identify abilities that appear generally beneficial, but this is easiest and least controversial in the case of abilities whose lack seriously limits our options in life. If I am positively stupid, hideous to look at, or constantly depressed, I may have a problem, but this has got little or nothing to do with the fact that you are more intelligent, better-looking, or generally more upbeat than I. The problem is not comparative in nature. If I am merely less intelligent or less handsome than you, I may not be able to reach the same professional heights as you or to find a partner who is physically as attractive as yours, but if my intelligence and my looks don't fall too much under the average then there is no reason to assume that I cannot also have a rich, fulfilling, and professionally successful life, and that is all I need and all I can reasonably want from life. What matters is not how much smarter you are (in comparison

[9] Peter van Inwagen (2001) has argued that even if somebody has a genetic disposition to a certain kind of antisocial or criminal behaviour (e.g. rape), this fact does not diminish their responsibility for committing the crime as long as they could have acted otherwise. In other words, only if the behaviour in question is genetically determined (rather than just genetically influenced) should we regard this as a mitigating circumstance. However, even if van Inwagen is right, my point here is not about moral responsibility but about the fairness or unfairness of having a propensity to a certain kind of behaviour. Even though it might be considered more unfair that I am bad and you are not if I just can't help being bad, it is arguably still unfair if being good is just much harder for me than for you.

with me) but merely how smart I am, and whether I am smart enough to fit in and to be able to live a good life. The same holds for all other natural assets. It is good to have some of them or all of them to some reasonable degree, but I don't have to be the best in everything to live a good life. Although with my limited abilities there are a lot of things in life that I cannot do, or cannot achieve, it would still be impossible to do most of those things even if I possessed all possible abilities to the highest possible degree. True, with my natural assets I could never hope to be a David Beckham, a Tiger Woods, or many others who excel in a particular area, but I could never be all of them anyway, especially not if everyone else were a natural-born Beckham or Woods.[10] As long as there is something I can do, something I am good at, and nothing that prevents me from fully participating in the world and society in which I live, I should be just fine. Why should we all want to be good at everything anyway? That is why Buchanan and colleagues very wisely concluded that 'at least for the foreseeable future (if not forever), the appropriate objective […] may be something more like the attainment of a "genetic decent minimum" […] than the elimination of all inequalities in natural assets' (Buchanan et al. 2000, p. 82). With this I can wholeheartedly agree.

The idea that we should all have exactly the same starting conditions to guarantee that we also have more or less the same opportunities in life strikes me as wrong-headed from the outset. It frames life as a competition and then seeks to eliminate all differences that might give anyone any advantage over anyone else, which makes any meaningful competition impossible.[11] We do, of course, compete with other humans, and we do compete more with them than we compete with other, non-human, species. That may be the reason we tend to ignore other animals and their abilities when we consider what is fair and what is not. But there is nothing really to be gained by levelling the genetic playing field for all humans so that nobody starts the competition with a disadvantage. Equality, even equality of opportunity, does not have any obvious intrinsic value. In fact, it does fairly obviously not have any intrinsic value. If we made the smarter ones dumber, the happier ones less happy and the prettier ones less pretty, then we would do as much for equality of opportunity than if we did the opposite. But nobody suggests that, and with good reason.[12] Life is not a race, and ethicists should not have to suppose that it is. We don't necessarily compete with each other. And even if we did (or where we do), it couldn't work if we were all the same. If the resources are

[10] And if everyone were, then we would certainly not value their talents much anymore. What we do value is the exceptional, not the commonplace.

[11] As Allen Buchanan (1995, p. 113) has pointed out, it would also expand the domain of moral responsibility to tyrannical dimensions.

[12] It might be thought that with regard to social assets we do in fact occasionally pursue a levelling-down approach to equal opportunity—for instance, when we implement progressive taxation systems that require the rich to contribute a larger percentage of their income than that which the poor are being asked to pay. However, the reason for this is not that we want to create equal opportunities or make sure that nobody is disadvantaged (and nobody advantaged) but simply that we assume, with good reason, that the rich can afford to pay a larger percentage of their income than the poor. We don't want to take away their advantage; we just think, in accordance with Rawls's difference principle, that everyone should benefit from their having more than the rest.

limited, then we can only achieve equality by holding people back. If we don't want to do that, if we do accept inequality at some stage, then we will, by guaranteeing the equality of the starting conditions, just postpone the time when inequality will materialize, which seems to defeat the purpose of the whole enterprise.

Now, it is commonly assumed that while it would be unfair not to grant everyone the same starting conditions, it is not unfair that people achieve different things in life through their personal virtues: their determination and dedication, the effort and work they put in to accomplish something, and so forth. But, of course, all this also requires certain abilities that are as much based on our genetic code as our intelligence, our longevity, or our propensity to develop cancer.[13] If it is unfair that I am less intelligent than you, then it is also unfair that I am lazier than you, less committed, less dependable, less trustworthy, and so on. So we would have to equalize these abilities too in order to *really* level the playing field, and once we have done that the only differences that would remain, if any, would be down to brute luck. So ironically, by levelling the playing field completely, in order to annihilate morally arbitrary distinctions, we would create a situation in which *all* differences would become morally arbitrary. That is because if there still are differences in people's situation, they could no longer result from differences in intelligence, talent, or virtue since all those differences have already been eradicated. Our whole life path would then be a lottery: entirely fair and at the same time entirely arbitrary.

Acknowledgments

I am grateful to Elizavet Kolovou for providing me with valuable information for the preparation of this chapter, and to John Ashton, Steve Clarke, Mark Coeckelbergh, Michele Loi, Kyle McNease, Inmaculada de Melo-Martín, Thomas D. Philbeck, Sagar Sanyal, Maartje Schermer, Edward Skidelsky, and Andres Vaccari for reading a first draft and providing critical feedback that helped me refine the argument that I was trying to make.

References

Bostrom, Nick (2004). 'Golden', <http://www.nickbostrom.com/fable/retriever.html>.

Bostrom, Nick (2005). 'The Fable of the Dragon Tyrant', *Journal of Medical Ethics* 31(5): 273–7.

Buchanan, Allen (1995). 'Equal Opportunity and Genetic Intervention', *Social Philosophy and Policy Foundation* 12(2): 105–35.

Buchanan, A., Brocks, D. W., Daniels, N., and Wikler, D. (2000). *From Chance to Choice: Genetics & Justice*. Cambridge: Cambridge University Press.

Chan, Sarah (2009). 'Should We Enhance Animals?', *Journal of Medical Ethics* 35: 678–83.

[13] This does not mean, of course, that all our abilities and talents are to an equal extent determined by our genes. Rather, my point is that there is *no* ability that is fully mine in the sense that it doesn't rely for its existence on certain aspects of my given nature, which I did not choose and which I did nothing to deserve.

Dvorsky, George (2008). 'All Together Now: Developmental and Ethical Considerations for Biologically Uplifting Nonhuman Animals', *Journal of Evolution & Technology* 18(1): 129–42.

Harris, John (2007). *Enhancing Evolution*. Princeton, NJ and Oxford: Princeton University Press.

Hauskeller, Michael (2013). *Better Humans? Understanding the Enhancement Project*. London: Routledge.

Holtug, Nils (1999). 'Does Justice Require Genetic Enhancements?', *Journal of Medical Ethics* 25: 137–43.

Hooker, Brad (2005). 'Fairness', *Ethical Theory and Moral Practice* 8: 329–52.

Lewens, Tim (2009). 'What Are 'Natural Inequalities'?' *The Philosophical Quarterly* 60(239): 264–85.

Mizrahi, Moti (2014). 'The Problem of Natural Inequality: a New Problem of Evil', *Philosophia* 42: 127–36.

Parker, Lisa (2012). 'In Sport and Social Justice, Is Genetic Enhancement a Game Changer?', *Health Care Analysis* 20: 328–46.

Persson, I. and Savulescu, J. (2008). 'The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity', *Journal of Applied Philosophy* 25(3): 162–77.

Rawls, John (1999). *A Theory of Justice*, rev. edn. Cambridge, MA: Belknap Press.

Savulescu, Julian (2005). 'New Breeds of Humans: The Moral Obligation to Enhance', *Reproductive BioMedicine Online* 10, supplement 1: 36–9.

Savulescu, J., Foddy, B., and Clayton, M. (2004). 'Why We Should Allow Performance Enhancing Drugs in Sport', *British Journal of Sports Medicine* 38: 666–70.

Tännsjö, Torbjörn (2009). 'Medical Enhancement and the Ethos of Elite Sport', in Julian Savulescu and Nick Bostrom (eds), *Human Enhancement*. Oxford: Oxford University Press, pp. 315–26.

van Inwagen, Peter (2001). 'Genes, Statistics, and Desert', in David Wasserman and Robert Wachbroit (eds), *Genetics and Criminal Behavior*. Cambridge: Cambridge University Press, pp. 225–42.

Veenhoven, R. and Choi, Y. (2012). 'Does Intelligence Boost Happiness?', *International Journal of Happiness and Development* 1(1): 5–27.

Walker, M. A. (2009). 'The Case for Happy-People Pills', *Free Inquiry* 29(5): 33–6.

15

Buchanan and the Conservative Argument against Human Enhancement from Biological and Social Harmony

Steve Clarke

15.1 Introduction

In all likelihood it is going to become possible for us to use drug therapies, genetic interventions, and other medical procedures, including procedures that augment human bodies with mechanical extensions, to increase the physical and mental capacities of many of us to above the normal upper limits for our species (Bostrom and Savulescu 2009). In other words, we will be able to 'enhance' ourselves. While some scholars welcome the prospect of human enhancement (e.g. Harris 2007; Savulescu 2006), many others view it with alarm; and it has attracted widespread opposition, especially (but not only) from conservative scholars. Prominent conservative opponents of human enhancement include Leon Kass (1997; 2003) and Francis Fukuyama (2002). In his recent *Beyond Humanity? The Ethics of Biomedical Enhancement* (2011), Allen Buchanan takes issue with a slew of prominent conservative arguments against human enhancement. For the most part I find his rebuttal of these arguments very persuasive. In this chapter I discuss one line of conservative argument against human enhancement that Buchanan considers briefly and dismisses quickly. This is the 'argument from biological and social harmony' (Buchanan 2011, pp. 161–2). In my view there is a variant of this line of argument that has much more going for it than Buchanan allows. I will not argue that it is strong enough to warrant the outright banning of human enhancement technologies, but I will argue that it gives us reason to be cautious about the widespread adoption of some possible human enhancements.

I will proceed as follows. In Section 15.2 I discuss Buchanan's rebuttal of conservative arguments against human enhancement, and describe the argument from biological

and social harmony in some detail, along with Buchanan's reasons for opposing it. In Section 15.3 I look at modern political history to develop a case for the importance of one version of the argument from biological and social harmony, and in Section 15.4 I consider evidence from biological anthropology that buttresses this argument. I then reconsider the argument from biological and social harmony, in light of the evidence discussed, in Section 15.5, before making my concluding remarks. I go on to conclude that there is a variant of the argument which should be taken much more seriously than Buchanan allows. I also conclude that there is nothing distinctively conservative about it. It is an argument that should appeal to liberals as much as it does to conservatives.

15.2 Buchanan on Conservative Opposition to Human Enhancement

Although Buchanan recognizes that many conservatives are conservatives because of their religious convictions, he devotes the bulk of his attention to secular arguments for conservatism in the tradition of Edmund Burke (1729–97) (Buchanan 2011, p. 146). He is right to do so. The secular arguments in this tradition are the conservative arguments with the widest potential appeal. Religious arguments for conservatism can only succeed in persuading those who share key tenets of the particular religious tradition in which they have been framed. Secular arguments for conservatism have the potential to persuade everyone, regardless of whether they are secular or religious. And the Burkean tradition is the most developed secular conservative intellectual tradition.

The key lines of argument against human enhancement that Buchanan considers build on traditional conservative scepticism about the possibility of effective social reform. Conservatives have long argued that ordinary fallible humans are subject to 'permanent constraints' which undermine deliberate attempts at social reform. Our cognitive capacities are limited and rendered unreliable by the influence of many of our motives and emotions. We may be inclined to work towards the betterment of our communities, or the betterment of all of humanity, but we also have selfish motives that undercut many of our pro-social tendencies and we experience emotions that can undermine our attempts to set aside selfish motives. People are not likely to go along with, for example, socialist schemes to develop fair, equitable societies, both because they are motivated by self-interest and not just the interests of the society in which they reside, and because their feelings of envy, pride, spite, and so on prevent them from suppressing selfish motives and acting in the interests of their society, even when they are persuaded that they ought to do so, or so say many conservatives (Buchanan 2011, p. 147). A closely related line of conservative reasoning, also directed against the possibility of effective social reform, has it that the sheer complexity of human society, coupled with our own cognitive limitations, ensures that deliberate attempts to improve human society are almost guaranteed to fail, and are

likely to cause more harm than good. Buchanan refers to this line of reasoning as the 'back-fire view' (Buchanan 2011, p. 146).

An obvious way to try to address conservative concerns about the consequences of our having the cognitive and affective limitations that we have is to attempt to use enhancement technologies to overcome these limitations. Cognitive and affective limitations may well be part of our nature, as conservatives assert, but crucial for the conservative argument against the possibility of social reform is that these are *permanent* features of our psychology. But why should we accept that they are permanent? Couldn't enhancement technologies be used to make us less selfish, less prone to be led astray by our emotions, and capable of more sophisticated cognition? Conservatives usually concede that it is at least possible that we could overcome the cognitive and affective limitations bestowed on us by nature. However, they usually also argue that our chances of doing so are remote. Human nature is exceedingly complex and our understanding of it is minimal. According to conservative critics of human enhancement, it is very likely that attempts to improve human nature will back-fire.[1]

Buchanan is unimpressed by conservative arguments against the possibility of effective human enhancement that invoke either the 'permanent constraint view' or the closely related back-fire view. He tells us that conservatives who appeal to these views place themselves in an 'uncomfortable position' (2011, p. 150). They need to show that we have a reasonably powerful set of cognitive capacities—powerful enough to enable us to know what our cognitive and affective abilities are, and to reliably discern their limitations. However, they also need to demonstrate that our cognitive capacities are insufficient to enable us to overcome those limitations (Buchanan 2011, p. 150). Conservative critics of human enhancement appear to assume that the level of cognitive capacity required to create reliable and safe human enhancements is greater than the level of cognitive capacity needed to reliably discern our cognitive and affective limitations. But this assumption seems unwarranted. The science behind human enhancement is in its infancy, and it is much too early to determine how much cognitive capacity is required to create reliable and safe human enhancements. It may turn out, contra conservative assumptions, that our current cognitive capacities are sufficient to reliably enable improvements to our cognitive and affective abilities.

Buchanan characterizes the argument from biological and social harmony as the argument that 'human biology and social institutions are in complex, balanced harmony, so that efforts to change the former may dangerously disrupt the latter' (Buchanan 2011, p. 161). It can be understood as a specific version of the more general conservative line of argument, which appeals to the back-fire view, in this case

[1] An illustration of such back-firing is contained in Antony Burgess's famous novel *A Clockwork Orange* (Burgess 1962). In it, Alex, a sociopathic and violent teenage delinquent, is given a (fictional) combination of aversion and drug therapy. The therapeutically reformed Alex finds himself unable to act on his sociopathic desires, and incapable of disobeying authority figures or defending himself. Burgess intends the reader to be at least as disturbed by the reformed Alex as by the sociopathic delinquent whom they first encounter at the beginning of the book.

appealing to social back-firing that results from biological intervention. Buchanan concedes that that there is 'a grain of truth' in the argument, which is that 'the social institutions we have are connected to our biology in ways we do not fully understand' (2011, p. 161). But he is no more impressed by it than he is by the more general conservative argument against the possibility of effective human enhancement. According to him, 'where the argument goes wrong is in assuming that there is a fragile, stable, harmony between social institutions and human biology' (2011, p. 161). He construes the relationship between social institutions and human biology as being far less liable to break down as a result of biological interventions than conservatives often suppose. He considers five ways in which we might alter our biology, and he assures us that there is no good reason for concern about the influence of any of these possible interventions on our social institutions. According to him,

> There is no good reason to think that remedying our inability to biosynthesize vitamin C or enhancing our immune systems or supplementing mechanical joint replacement technology with regenerative stem cell research technology or extending the human life span by 20 years or increasing average IQ by 10 points would cause unacceptable social disruption. (2011, p. 161)

In my view, conservatives are right to be concerned about the fragility of at least some relations between our biology and our social institutions. Some of the social institutions that contemporary conservatives (and others) most care about are liable to collapse under the influence of a range of interventions, including biological ones. I also think that two of Buchanan's examples of apparently harmless biological interventions are actually examples of interventions that might well destabilize social institutions we care about. I'll approach the task of explaining why I hold these views indirectly, starting by examining an aspect of conservative thought that Buchanan (2011) does not focus on.

15.3 Conservatism and Free Societies

According to Kekes, 'The source of conservatism is a natural attitude that combines the enjoyment of something valued with the fear of losing it' (1998, p. 5). Like other conservatives, Thomas Hobbes (1651[1962]) valued social stability, and he was convinced that it would soon be lost without the stabilizing influence of a monarchical government. According to him, life without social stability is of very little value. One thing that a government needs to do to provide ongoing social stability, according to conservatives, is to preserve the traditions and institutions that the society it governs is built around (Kekes 1998, pp. 36–41; Scruton 2001, pp. 30–6). To the extent that life in our current society is stable and enjoyable, it is because preceding members of our society have developed traditions and institutions that enable that stability and enjoyment. If these are abandoned, or if they collapse, then in all likelihood our lives will become significantly less stable, and significantly less enjoyable, than they are now.

The political traditions and institutions that have developed in liberal Western countries over the last couple of centuries are democratic ones. The residents of these countries have found a way of living that they have come to value more highly than the ways their forebears lived, which were under the rule of monarchs. Life in a liberal democracy enables people to enjoy opportunities that they would be unlikely to enjoy under a monarchical government, including the opportunity to criticize their government publicly, and the opportunity to try to replace it without resorting to violence, through democratic processes. Recent and contemporary Western conservatives value democratic traditions and institutions, placing a very high value on the freedom from tyranny that democracy enables (Haidt 2012, pp. 175–6). Consequently, they have more to fear than their seventeenth-century counterparts, such as Hobbes. Like Hobbes, they have the destabilization of the social order and the chaos which would ensue to fear. Unlike him, they also have the collapse of democracy and a reversion to the oppressive rule of the many by the few to fear.

Recent and contemporary conservatives differ from liberals and libertarians, who also value life in free democratic societies, because *inter alia*, conservatives stress the difficulty of achieving and maintaining the traditions and institutions that enable free democratic societies to function. According to mainstream conservatives, many of us will act in selfish, foolish, and capricious ways if we are allowed to, and we may jeopardize the future of the traditions and institutions that enable free democratic societies to function while behaving in these ways.[2] Conservatives hold that restrictions on selfish, foolish, and capricious activity are often necessary to prevent society from collapsing or reverting to tyranny. Unlike liberals and libertarians, contemporary conservatives have few qualms about restricting individual freedoms to prevent behaviour that threatens the future of democratic traditions and institutions.

Modern liberal democracies are relatively recent forms of government, emerging only in the last 250 years, after thousands of years of human civilization in which monarchy and other forms of dictatorship were omnipresent. While some liberal democracies, such as the US and the UK have proved to be stable, many others have not. The First French Republic lapsed into dictatorship after only 12 years (1792–1804). The Second French Republic was even shorter lived, lapsing into dictatorship after less than 4 years (1848–52). Many other attempts to establish democratic societies also faltered in their early stages. Famous twentieth-century examples include the Second Spanish Republic, which lasted 8 years (1931–9), and the Weimar Republic in Germany, which lasted for 11 years (1919–30). Contemporary democratic multiparty states last 17.5 years on average. This is more than the average contemporary military dictatorship, which lasts a mere 11.1 years, but considerably less than contemporary monarchies, which last an average of 25.4 years.[3]

[2] See, for example, Kekes 1998, pp. 13–15.

[3] See Hadenius and Teorell 2007, pp. 149–52. Their data is from the years 1972–2003. Stier and Mundt (1997) examined the stability of democracies between 1776 and 1992. According to them the average democratic regime lasted either 10.45 or 22.47 years (depending on whether or not data about regimes that

We don't really understand why some democratic states are long-lasting while others are not.[4] Successful systems of democratic government, which provide the checks and balances needed to enable free democratic governments to survive over the long term, such as the Westminster system in the UK, Canada, Australia, and New Zealand, have emerged slowly, through trial and error, rather than as a result of careful, deliberate planning (Rhodes et al. 2009). The fact that successful democratic institutions and traditions have developed slowly over time is no surprise to conservatives who emphasize the propensity of deliberate social planning to back-fire. Some prominent 'neoconservatives' of the early twenty-first century, such as the former US Secretary of Defense Donald Rumsfeld, supposed that they understood how to create liberal democracies from first principles. Under their influence, George W. Bush's administration attempted to impose liberal democratic institutions on Iraq and Afghanistan. These attempts are generally regarded as failures and they have served to strengthen the mainstream conservative conviction that deliberate social planning is almost guaranteed to fail (Fukuyama 2006, pp. 29–31).

Conservative fears that free democratic societies are liable to lapse into dictatorship seem well placed. Dictatorship, or some variant of it, has been the usual form of government in human societies for thousands of years, and significantly many free democratic societies have not proved stable and have lapsed into dictatorship. In this context, we ought to be very wary of any changes that impact on the social order in democratic societies, including biological ones. Biological changes that influence the behaviour of (sufficiently many) individuals will, almost invariably, cause changes to the social structures in our society, although exactly what their effects will be and exactly how far they will extend is difficult to predict.

15.4 Conservatism and Biological Anthropology

The claim that biology has a significant influence on social organization in at least some species is undeniable. The various eusocial insects exhibit very high levels of social organization, which is a consequence of biological causes (Hölldobler and Wilson 2008). This causal relationship is particularly apparent among insects with haplodiploid sex-determination systems, such as the hymenoptera (an order of insects that includes ants, bees, and wasps), which are all eusocial. Female haplodiploid insects share three-quarters of their genetic material with their sisters but only half with their mothers. Haplodipoid insects can more effectively replicate their own genetic material by helping to rear sisters than by reproducing directly. Large numbers of eusocial female haplodiploid insects, such as 'worker' ants, forego opportunities to reproduce

either began before 1776 or continued to exist after 1992 are included in the data set from which averages are determined).

[4] There is a burgeoning body of empirical work on the causes of stability and instability of democracy. For a recent example of this literature, which underscores the complex interplay of causes that contribute to democratic stability and instability, see Diskin, Diskin, and Hazan 2005.

and instead devote their energies to ensuring that 'queen' ants, who are their mothers, are able to reproduce sisters for them at a maximal rate.[5]

We humans are both genetically and socially different from eusocial insects. However, it should not be assumed that our biology is unconnected to the modes of social organization we tend to adopt, in advance of consideration of the relevant evidence. One way to try to understand human social behaviour, and to try to figure out what factors cause it to vary, and in what ways, is to look at the social behaviour of the species to which we are most closely related: chimpanzees, gorillas, and bonobos. The members of all of these species live in social groups characterized by dominance hierarchies (Mameli 2013, p. 920).[6] Also, in all of these species, recurring attempts by subordinates to overthrow dominance hierarchies and replace them with new dominance hierarchies headed up by usurpers can be observed. These attempts are usually kept at bay by dominant individuals, with displays of aggression and acts of violence (Knauft 1991, p. 396; Boehm 1999, pp. 130–7).[7]

Humans, chimpanzees, gorillas, and bonobos are all descended from a common ancestor species. Humans split off from the ancestor species perhaps 5 million years ago. The ancestor species almost certainly lived in hierarchical social groups, as did early humans (Boehm 2012, p. 151). By the Late Pleistocene (126,000–11,700 years ago), humans had undergone a remarkable social transformation and had begun to live in groups that were egalitarian, in the sense that they lacked hierarchies and shared most of their important resources equally (Mameli 2013, p. 919). Hunter-gatherer societies, which were the only form of human society until 8,000 years ago, are exclusively egalitarian, in the above sense of the term 'egalitarian'. Insofar as they have leaders, they have ones who are able to secure the allegiance of other members of their society by representing their collective wishes, rather than by imposing their will on others. It appears that hunter-gatherers did not lose the tendency to try to dominate, which, it seems reasonable to believe, was characteristic of the ancestor species. However, at some point, hunter-gatherer communities began to exhibit 'anti-big man behaviour', which became universal (Smith et al. 2007, p. 286). The policing of bullies—who would dominate other hunter-gatherers if they could—is a form of behaviour that has been observed in all known hunter-gatherer communities. Such policing is usually persistent and aggressive, and can involve capital punishment (Boehm 1999, pp. 72–84, 2012, pp. 83–7).

[5] Some eusocial worker insects attempt to reproduce directly, as well as helping to enable their queen to reproduce. In many eusocial insect colonies, this sort of behaviour is monitored, and attempts are made to prevent it, by eating or removing eggs laid by colony members other than the queen. This phenomenon is known as 'worker policing'.

[6] Gorillas and chimpanzee have male dominance hierarchies. Bonobo societies have dual dominance hierarchies, with a female hierarchy existing alongside a male one (Wrangham and Peterson 1996, p. 205).

[7] Bonobos differ from chimpanzees and gorillas in that—like humans—they seem to have the ability to mediate at least some conflicts (Knauft 1991, pp. 396–7). There are ongoing debates about the effectiveness of bonobo conflict-mediation techniques. See De Waal 2013 for a recent discussion.

With the rise, in the last 8,000 years, of settled communities based on agriculture, humans overwhelmingly reverted to living in hierarchical societies dominated by chiefs. These chiefs were able to overcome the effects of anti-big man behaviour and were able to impose their will on others (Diamond 2012, p. 17).[8] Structured hierarchies, such as hereditary monarchies, continued to be characteristic of human societies through the early phases of industrialization, until the last 250 years, when we began to see the emergence of liberal democratic governments.[9] Democracy is a recent form of egalitarian social organization.[10] Unlike the other great apes, humans have the capacity to live in both hierarchical and egalitarian societies, and they can switch between these two modes.

Why did hunter-gatherers become egalitarian in the first place? Boehm suggests that this shift may have resulted from the development of early weapons, such as the spear (1999, pp. 177–81).[11] When weapons became widely available, a shift in the balance of power between dominant alpha males and their subordinates took place. It became easy for armed subordinates to kill alpha males, especially at times when alpha males were unarmed, such as when they were sleeping. So it became harder for alpha males to dominate others, who were usually disinclined to be dominated, than it had been before the development of early weapons.[12] Hunter-gatherer egalitarianism can be understood as a form of compromise between would-be leaders, who all realize that they have the power to overthrow any other leader, but lack the power to then go on to dominate other members of their community.

A reversal of the shift in the balance of power that led to hunter-gatherer egalitarianism took place when and where settled agricultural societies succeeded hunter-gatherer societies. It remained easy for alpha males to be killed with weapons in settled agricultural societies, but it also became easier for alpha males to form a privileged caste within a settled agricultural society and to coordinate their activities to protect a leader from assassination, and prevent rebellions by subordinates. Members of privileged castes in settled agricultural societies tended to remain loyal to their leaders

[8] The relationship between the creation of hierarchies and the development of permanent human civilization is a theme that is often stressed by anarchist and socialist authors. See, for example, Bookchin 1982, pp. 3–8.

[9] I do not mean to imply that there were no democratic governments before the rise of liberal democracy in the last 250 years. Of course, there were occasional early democracies, most notably Athens in the sixth and fifth centuries BC, but democracy was extremely uncommon until the last 250 years.

[10] Most modern democracies make some form of commitment to universal human rights. The egalitarianism characteristic of hunter-gather societies typically did not involve any recognizably homologous intellectual commitment. Hunter-gatherer egalitarian values are typically not universal. They typically only concern relations with in-group memberss and do not concern relations with out-group members. Thanks to Russell Powell for this point.

[11] Boehm suggests that this may have happened by 500,000 years ago (1999, p. 181).

[12] Boehm (2012) offers a slightly different, but complementary, explanation of the rise of egalitarian values among hunter-gatherers. This is that hunter-gatherer groups with egalitarian values outcompeted groups that lacked egalitarian values, due to the enhanced ability of members of groups with shared egalitarian values to cooperate during big-game hunting and to distribute meat gained by hunting efficiently. See also Mameli 2013, who refers to Boehm's explanation as the view that 'meat made us moral'.

because it was usually in their interests to do so. If a leader were to be overthrown then the privileged future of the members of the caste who had been loyal to that leader would be jeopardized.

The recent rise of democracy seems to have been driven, in part at least, by the high value that individuals usually place on freedom from domination by others. Thousands of years of living in dominance hierarchies has not quashed the human desire to avoid domination.[13] The acceptance of domination by tyrants brings with it a combination of costs and benefits. Domination by a tyrant usually brings with it the benefit of ongoing social stability; and the realization of many of our preferences requires the cooperation of other members of our society, which is enabled by ongoing social stability. But a tyrant is also likely to suppress the expression of many of our preferences, particularly those that might lead to threats to their rule. Furthermore, the rule of tyrants is often capricious, and the whims of tyrants can sometimes be a source of (usually mild) social instability. Democratic government promises life with the benefits of ongoing social stability but without the costs associated with tyranny.

15.5 The Argument from Biological and Social Harmony Reconsidered

Recent and contemporary Burkean conservatives value life in free democratic societies and fear that they could be deprived of this good. They have good reasons to fear that they may indeed be so deprived. The evidence we have considered, from modern history and from biological anthropology, suggests that the social institutions of free egalitarian human societies, including modern democracies, are not particularly robust and are liable to be replaced by the institutions of unfree hierarchical societies. We understand some of the causal factors that lead to the collapse of egalitarian societies and to their replacement by tyrannical hierarchies. We understand, for example, that there are some individuals who crave power over others and will use force if they can to obtain and then retain such power. However, we are a long way from understanding the complicated mix of causal factors that lead free egalitarian human societies to collapse into tyrannies.

Bearing in mind our uncertainty about the causes of egalitarian societies collapsing and turning into tyrannies, as well as our reasons for thinking that these societies are liable to do so, let us ask ourselves whether there are good reasons to fear that the widespread use of some human enhancements might lead to such a result. In general, those of us who value democracy should be very wary of changes, including changes to our

[13] We may have an evolved preference for freedom from tyranny. In his recent *The Righteous Mind*, Jon Haidt speculates that a key component of contemporary morality, the sense that liberty is to be valued and oppression to be disvalued, evolved in humans over the many thousand years in which we lived exclusively in egalitarian hunter-gatherer communities. This may have resulted in the development of a 'moral module' in the brain that instinctively values liberty and rejects oppression (2012, pp. 172–4).

biology that could lead to shifts in the relative balance of power between groups within society. Democratic societies can, and do, tolerate some imbalances of power between competing groups. However, such shifts can enable newly empowered groups to dominate other members of society; and when they find themselves in dominant positions, they may seek to ensure that they can continue to dominate by doing away with democratic checks and balances on their power.[14] Can we conceive of ways in which the use of particular forms of enhancement might lead to significant shifts in the relative balance of power between groups within society? Let us consider Buchanan's five examples of allegedly harmless enhancements:

- imbuing people with the ability to biosynthesize vitamin C
- enhancing our immune system
- supplementing mechanical joint replacement technology with regenerative stem cell research technology
- extending lifespans by 20 years
- increasing average IQs by 10 points

The first three examples seem harmless enough. I can think of no reason why imbuing people with the ability to biosynthesize vitamin C would lead to shifts in the relative balance of power between groups within society, assuming, of course, that access to this postulated enhancement was reasonably equitable. The same is true of enhancing our immune system and developing stem cell technologies. The introduction of these enhancements would be unlikely to disrupt the current social order, provided that the benefits that they yield are distributed in a reasonably equitable manner.

However, Buchanan's other two examples are more concerning. Extending the average lifespan by 20 years would result in a shift in the balance of power between different age groups. A greater proportion of our societies would be middle-aged and elderly and, all things being equal, the middle-aged and elderly would hold more power than they currently have and the young would be dispossessed of some power. It is not clear that postulated future societies enhanced in this way would remain free and democratic. At any rate, there seem to be reasons to fear that they would not. Newly empowered groups within a society that represented the interests of the middle-aged and/or the elderly might seek to consolidate their new-found power, impose absolute rule over their society, and overthrow democracy. Another possibility is that groups that represent the interests of the young might resent their newly disempowered status and seek to overthrow democratic governments. Either way, democracy would be faced with new threats.

[14] Authors working in the political pluralist tradition, such as Dahl (2005), stress the ability of well-formed democratic governments to manage shifts in power between competing interest groups within society and so ensure ongoing social stability. This may be an ideal that democracies should strive for, but judging by the rate at which contemporary democracies collapse, it is apparent that it is often not met. We do not currently know how to structure democratic governments so as to ensure that this ideal is met.

If we raised the average IQ by 10 points we would increase the abilities of those within postulated future enhanced societies, who seek to overthrow those societies, to formulate effective plans to achieve that goal. We would also increase the intellectual abilities of those who aim to prevent the overthrow of free democratic societies, so it might be presumed that an overall balance of power would be maintained. However, it is far from clear that a balance of power would be maintained. If Boehm's aforementioned suggestion (1999, pp. 177–81) is correct, the invention of early weapons resulted in a shift in the balance of power within hunter-gatherer societies, even in situations when and where access to those weapons was evenly distributed. It proved harder for armed alpha males to persistently oppress armed subordinates than it is was for unarmed alpha males to persistently oppress unarmed subordinates. Plausibly, it requires more intellectual capacity to maintain the social order in liberal democratic societies than it does to overthrow that social order and replace it with a dictatorship; so an evenly distributed increase in IQ might well result in a shift in the balance of power. It could make groups who seek to overthrow democracy significantly more effective while only enabling less significant improvements to the efficacy of those who seek to maintain the social order in liberal democratic societies. It would thereby result in a power shift in favour of those who would seek to overthrow democracy.[15]

Perhaps aware that, like me, some of his readers will be concerned by the social disruption that extending average lifespans and increasing average IQs would cause, Buchanan offers a line of argument that is intended to allay our fears. According to him, 'It is worth recalling that human life expectancy has risen far more than that [20 years] since 1900 and that there has been a significant rise in average IQ over the past 50 years—the so-called Flynn effect—but this has not resulted in the social fabric unraveling' (2011, pp. 161–2). I don't find these words reassuring and nor should others. The twentieth century is not usually seen as a particularly stable period in human history.[16] It involved two world wars, the rise and decline of communism, and—as discussed already—the rise and fall of various free democratic governments. We don't really understand why these various events occurred. It may be that shifts in the relative balance of power between competing groups in society, caused by the extension of lifespans and the rise in average IQ, played a part in causing them. I'm not arguing that the extension of lifespans and the rise in average IQ definitely played these causal roles. Rather, I'm arguing that given what we know

[15] Pinker (2011, pp. 664–5) identifies a correlation between increasing IQ and a propensity for democratic governance. It could be that, far from being a threat, increases in IQ are responsible for entrenching democratic values. Currently we don't know one way or the other.

[16] Pinker (2011, pp. 189–294) argues that, despite appearances, we should understand the twentieth century as being relatively stable and peaceful. While Pinker argues, persuasively I think, that violence has declined over the course of human history, and that the last century was less violent than earlier centuries, it does not follow that the twentieth century was more stable than earlier centuries. In my view it was less stable than many of the centuries that preceded it, which were not disrupted by global-scale wars and by any global-scale ideological conflicts as pervasive as the conflict between the Western Bloc and the Eastern Bloc during the 1947–91 Cold War.

and what we don't know about the causes of instability in the twentieth century, there is sufficient reason for us to be wary.

15.6 Concluding Remarks

I've argued that at least one variant of the conservative argument from biological and social harmony has more going for it than Buchanan allows. A human social institution that most of us living today desire—democratic government—is fragile. Contemporary democratic regimes last, on average, less than two decades (Hadenius and Teorell 2007), and they are easily replaced by a common form of human social institution that is much less desirable: dictatorship. The introduction of human enhancements that can cause significant shifts in the relative balance of power between competing groups in a society is liable to lead to the collapse of democratic states and their replacement by dictatorships, and this is an outcome that conservatives are right to fear. This variant of the argument from biological and social harmony should not be understood as an argument against enhancement *tout court*. The introduction of enhancements needs to be considered on a case-by-case basis. When there is no plausible reason to think that the introduction of a particular enhancement could lead to shifts in relative power between competing groups within society, and to the collapse of democracy, then this variant of the argument from biological and social harmony does not provide a reason to oppose the introduction of that enhancement.

This variant of the argument from biological and social harmony might be conjoined with an argument for a precautionary approach, when dealing with situations of uncertainty, to produce an argument against enhancement *tout court*. We might argue, for example, that because we do not know for sure that the enhancement of our ability to biosynthesize vitamin C will not threaten social stability, we should not take the risk of attempting to enhance that ability. I reject precautionary approaches to action under uncertainty (Clarke 2009), so I would not go along with such a conjoined argument. However, a proponent of the precautionary principle as a guide to action under uncertainty might well accept such a conjoined argument and take it to constitute an argument against enhancement *tout court*.

Another point worth noting is that this variant of the argument from biological and social harmony is best deployed against the mass use of particular enhancement technologies. For the most part, the one-off use of enhancement technologies is very unlikely to lead to significant social instability and to the onset of tyranny. There may be exceptions to this generalization. If one person were to increase their IQ by 10 points, this would be vanishingly unlikely to lead to them being able to figure out how to cause a liberal democracy to collapse and how to become a tyrant. However, if someone were to suddenly increase their IQ by 200 points, and become more intelligent than any human who has ever lived, then they might acquire enough cognitive capacity to be able to figure out how to cause a liberal democracy to collapse and how to

become a tyrant.[17] There are also possible cases that sit midway between the mass adoption of an enhancement and the adoption of it by one individual. The adoption of an enhancement by the members of a group within society, and not by members of the society who are outside of that group, is worrying if and when it leads to a shift in the balance of power between members of that group and other members of the society.

Buchanan has identified the argument from biological and social harmony as a conservative argument against enhancement (Buchanan 2011, p. 161). While it is a form of argument that has been employed by dyed-in-the-wool conservatives, such as Leon Kass, as Buchanan notes (2011, p. 161), there is nothing distinctively conservative about it. Burkean conservatives are not the only people who value living in free democratic societies and who fear that they may lose the enjoyment of this value. Placing a high value on liberty is definitional to liberalism. If liberals are persuaded that the introduction of particular enhancements involves a significant threat to the viability of current free democratic states, then liberals can also deploy a variant of the argument from biological and social harmony to oppose the introduction of those enhancements.[18]

References

Boehm, C. (1999). *Hierarchy in the Forest: the Evolution of Egalitarian Behaviour*. Cambridge MA: Harvard University Press.

Boehm, C. (2012). *Moral Origins: the Evolution of Virtue, Altruism, and Shame*. New York: Basic.

Bookchin, M. (1982). *The Ecology of Freedom: the Emergence and Dissolution of Hierarchy*. Palo Alto: Cheshire Books.

Bostrom, N. and Savulescu, N. (2009). 'Human Enhancement Ethics: the State of the Debate', in J. Savulescu and N. Bostrom (eds), *Human Enhancement*. Oxford: Oxford University Press, pp. 1–22.

Buchanan, A. (2011). *Beyond Humanity? The Ethics of Biomedical Enhancement*. Oxford: Oxford University Press.

Burgess, A. (1962). *A Clockwork Orange*. London: William Heinemann.

Clarke, S. (2009). 'New Technologies, Common Sense and the Paradoxical Precautionary Principle', in Paul Sollie and Marcus Duwell (eds), *Evaluating New Technologies: Methodological Problems for the Ethical Assessment of Technological Developments*. Dordrecht: Springer, pp. 159–73.

Dahl, R. A. (2005). *Who Governs? Democracy and Power in an American City*, 2nd edn. New Haven: Yale University Press.

[17] Similarly, some scholars are concerned that a rapid rise in machine intelligence, possibly as a result of 'the singularity' occurring, would lead to the end of liberal democracy and the creation of a new dictatorship of machines. For further discussion, see Yudkowsky 2008.

[18] Thanks to Alberto Giubilini, Russell Powell, and Julian Savulescu for helpful comments on earlier drafts of this chapter, as well as audiences at the workshop Human Enhancement: the Moral Challenge, held at the University of Melbourne in 2014, the Australasian Association of Philosophy Annual Conference 2014, and an audience at the Australian Catholic University.

De Waal, F. (2013). *The Bonobo and the Atheist*. New York: W. W. Norton.

Diamond, J. (2012). *The World until Yesterday*. New York: Viking.

Diskin, A., Diskin, H., and Hazan, R. Y. (2005). 'Why Democracies Collapse: the Reasons for Democratic Failure and Success', *International Political Science Review* 26(3): 291–309.

Fukuyama, F. (2002). *Our Posthuman Future*. New York: Farrar, Straus and Giroux.

Fukuyama, F. (2006). *America at the Crossroads: Democracy, Power, and the Neoconservative Legacy*. New Haven, CT: Yale University Press.

Hadenius, A. and Teorell, J. (2007). 'Pathways from Authoritarianism', *Journal of Democracy* 18(1): 143–57.

Haidt, J. (2012). *The Righteous Mind: Why Good People are Divided by Politics and Religion*. New York: Pantheon.

Harris, J. (2007). *Enhancing Evolution: the Ethical Case for Making Better People*. Princeton NJ: Princeton University Press.

Hobbes, T. (1651) [1962]. *Leviathan*. London: Collins/Fontana.

Hölldobler, B. and Wilson, E. O. (2008). *The Superorganism: The Beauty, Elegance and Strangeness of Insect Societies*. New York: W. W. Norton.

Kass, L. (1997). 'The Wisdom of Repugnance', *The New Republic*, 2 June, 17–26.

Kass, L. (2003). 'Ageless Bodies, Happy Souls: Biotechnology and the Pursuit of Perfection', *The New Atlantis*, Spring, 9–28.

Kekes, J. (1998). *A Case for Conservatism*. Ithaca, NY: Cornell University Press.

Knauft, B. M. (1991). 'Violence and Sociality in Human Evolution', *Current Anthropology* 32(4): 391–410.

Mameli, M. (2013). 'Meat Made us Moral: a Hypothesis on the Nature and Evolution of Moral Judgement', *Biology and Philosophy* 28: 903–31.

Pinker, S. (2011). *The Better Angels of Our Nature*. New York: Viking.

Rhodes, R. A. W., Wanna, John, and Weller, Patrick. (2009). *Comparing Westminster*. New York: Oxford University Press.

Savulescu, J. (2006). 'Justice, Fairness and Enhancement', *Annals of the New York Academy of Science* 1093: 321–38.

Scruton, R. (2001). *The Meaning of Conservatism*, 3rd edn. Basingstoke: Palgrave.

Smith, K. B., Larimer, C. W., Littvay, L., and Hibbing, J. R. (2007). 'Evolutionary Theory and Political Leadership: Why Certain People do not Trust Decision Makers', *The Journal of Politics* 69(2): 285–99.

Stier, M. and Mundt, Robert J. (1997). 'Is Democracy Stable? Compared to What? A Preliminary Exploration', paper presented at the American Political Science Association Annual Meeting, 1997, <http://www.stier.net/writing/demstab/stability.htm>, accessed 12 August 2014.

Wrangham, R. and Peterson, D. (1996). *Demonic Males: Apes and the Origins of Human Violence*. London: Bloomsbury.

Yudkowsky, E. (2008). 'Artificial Intelligence as a Positive and Negative Factor in Global Risk', in N. Bostrom and M. Circovic (eds), *Global Catastrophic Risks*. Oxford: Oxford University Press, pp. 308–45.

16

Moral Enhancement, Enhancement, and Sentiment

Gregory E. Kaebnick

16.1 Introduction

The debate about human enhancement has often been thought to pit hardliners against freethinkers. Those opposed to enhancement have sometimes come across as negative rather than as positive, as merely rejecting change rather than as honouring deep attachments to the world, as rigid rather than flexible, as maintaining a long levee against the flood of innovation, and as unwilling ever to admit that innovation might in fact be attractive. At any rate, this has been the view from the side of the debate in favour of enhancement. The picture was always, I think, a bit of a caricature, but it has become less accurate as the debate about enhancement has evolved. Part of what is helpful about Nicholas Agar's recent work on enhancement (2007, 2014) is that it integrates endorsement of and resistance to enhancement, explaining his resistance in fundamentally positive terms and working towards an attractive middling position. John Harris, too, now resists some forms of enhancement (for reasons discussed in Section 16.3) while strongly embracing other forms (2011). From the other side of the debate, Erik Parens has also moved towards more complex positions on enhancement (2005, 2009, 2015).

This chapter will try to further develop what Parens (2015) has called the 'second wave' of the enhancement debate, taking as a focal point the prospect of moral enhancement—that is, the enhancement of the capacities for deliberating on, settling on, and acting on moral judgements. It will argue, that is, that somebody who tends to resist enhancement might nonetheless be open-minded about moral enhancement. (The chapter will not offer a full-blooded argument for moral enhancement; it is the possibility of a middling position that is of most interest here.) Looked at one way, moral enhancement is a good place to look for open-mindedness: morality is so important to human well-being that enhancing it is likely to be particularly attractive. It is easy to accept unenhanced muscles (few people really need particularly good muscles), but it might be much harder to accept ordinary moral functioning.

Indeed, it's reasonable to ask whether the world can continue to abide unenhanced morality, given the enhanced capacities its denizens now have to be bad (Persson and Savulescu 2012).

Looked at another way, however, moral enhancement would be a surprising place to find open-mindedness about enhancement from those broadly resistant to it. The critic might seem to be resisting enhancement at the edges of life while accepting it at the core, insisting on ordinary muscles while smiling benignly on alterations that affect some of our deepest guiding ideas and values—our understanding of morality and freedom, human flourishing, and the positions on which the initial resistance to enhancement is based, such as our understanding of the human relationship to nature.

Whether moral enhancement would have these effects depends, however, on what it actually was and how we went about doing it. This chapter will explore several of these issues, seeking clarity on a series of questions about what moral enhancement would consist of, whether it would undermine morality and freedom, and whether it would threaten the enhancement sceptic's understanding of the desirable human relationship to nature.

16.2 What Would Moral Enhancement Be?

One point to note right at the outset of any serious, well-reasoned discussion of biomedical enhancement is that no serious, well-reasoned effort at biomedical moral enhancement is likely in the near future. Certainly there is interesting research going on regarding drugs that appear to affect aspects of human moral behaviour and into genes that appear to play a role in aspects of human moral behaviour, and it is just possible that some biomedical moral enhancement is already happening more or less incidentally in the course of treating human behavioural disorders, such as depression. But this work is just a bare beginning and might prove to be a false start. Serious moral enhancement might end up deploying a range of interventions that are not even yet on the table.

The lesson to be learned from the current research is not that moral enhancement is under way but that it appears in principle to be possible. The simple fact that human moral behaviour can be modified by biomedical as well as social means is enough to establish that, at some point, serious moral enhancement may be possible. There are physical attributes (genes, neurotransmitters, hormones, etc.) that give different kinds of animals different behavioural attributes—making some aggressive and others less aggressive, for example—and within any species there are physical attributes that give different individuals different temperaments. These attributes are many, working together in inordinately complex, mutually interacting systems with each other and with the organism's environment, but they are in principle comprehensible and modifiable. It is plausible, then, that we could identify behavioural attributes we think

are desirable and then identify and make physical changes that at least make those behavioural outcomes more likely.

Moreover, the fact that this capability probably lies in the far future does not seem to be a reason not to begin thinking about it. In fact, now is exactly when we should be thinking about it. (However, taking the long view could have an effect on some arguments for moral enhancement. If the argument is that we should accept enhancement because contemporary problems are insurmountable without it, then the right response may be, 'That's unfortunate.')

Supposing, then, that it might happen, what would it look like? Several different kinds of biomedical moral enhancement have been envisioned.

The simplest form imaginable, though it might exist only in the imagination, would consist of a technological trick that flatly assured morally good behaviour—good in the sense that it conformed to good moral judgements but not in the sense that it reflected the agent's good judgements since it might not be connected to agent's own judgements. Julian Savulescu, Thomas Douglas, and Ingmar Persson (2014) envision a scenario, set in the year 2100, of 'perfect mind control' in which a combination of neuroscience and information technology leads to the implantation of advanced iPhones that allow governments to monitor citizens, and instantaneously override bad intentions and install (as it were) good ones.

In addition to being the least plausible form of moral enhancement, this would also be the most objectionable. There are obvious and enormous technological challenges in reading, communicating, and changing a person's intentions, but even before these arise there is the difficulty of articulating them. An intention is a plan of action, but a given act can usually be described in a variety of quite different ways, no one of which is the single correct description (Douglas 2011). A given act might be described both as bringing a volume of iron quickly into contact with a human cranium and killing a person, and as desperately defending oneself with a frying pan against an armed intruder. The morally interesting descriptions are those with rich contextual and social detail, of course. The mind-control system would have to not only read the right descriptions but also apparently write the right descriptions, replete with rich information about the setting in which the action is to take place.

That is unlikely to happen. If it did, however, it would take over the business of thinking from the agent. (Would-be murderers would have the sensation of having changed their minds, although in fact the change had been effected externally.) In effect, moral agents would be deprived of moral agency, which might be a loss worse than that of the lives that would have been destroyed by the would-be murderer (Harris 2011). (I return to the relationship between enhancement and freedom in Section 16.3.)

There are also some strictly cognitive enhancements that could be seen as bringing about moral enhancements. By 'strictly cognitive' I refer to the capacity to reason, and I am excluding dispositional traits that have more to do with the emotions, such as sympathy. Insofar as moral judgements depend on capacities to, for example, engage in

self-reflection, hold many details in mind, do logic (as might be involved in recognizing inconsistencies between one's intentions and the intentions one expects others to act upon), and make cost–benefit estimations, then cognition is important to morality, and better cognition might therefore lead to moral enhancement as a follow-on effect.

These aspects of cognition capture quite a bit of what many philosophers have thought of as the heart of moral reasoning, and given the right starting valuations—a commitment to maximizing benefits or to acting only on those intentions that one wants others to act upon—enhancing them ought therefore to go a long way, from a philosophical perspective, towards achieving moral enhancement. Still, in and of themselves, cognitive enhancements do not necessarily produce moral enhancements, for one can be very intelligent yet quite wicked.

The converse of a technological trick that took over the agent's moral thinking would be a trick that forefronted the agent's own capacity for moral thinking. The social psychologist Joshua Greene (2013) argues that moral thinking tends to occur in two general forms: a fast version that relies on the emotional dispositions and a slow version that makes more use of thoughtful deliberation. The latter tends to go better, he argues. If so, then some moral enhancement might be achieved just by encouraging deliberativeness, and a little scepticism and humility.

A fourth strategy for moral enhancement would focus on the dispositional traits. These have to do with perception and motivation, and they are, of course, part of cognition under a broader and more interesting understanding of human cognition. They include 'pro-social attitudes' such as sympathy, trust, and generosity, which bear in different ways on how one responds to other people and their interests, self-sacrifice (thinking of that as a comparatively marked tendency to give less weight to one's own interests when confronted with the needs of others), and a relative lack of aggressiveness (understanding that as a tendency to be drawn towards violence or a threat of violence in order to secure one's own interests). Other relevant dispositions might be thought of as important enabling skills. Savulescu, Douglas, and Persson (2014) identify moral imagination (understanding this as the capacity to envision the emotional significance for oneself or for someone else of a given course of action) and impulse control as helpful targets for moral enhancement, for example.

16.3 Would Moral Enhancement Undermine Morality and Freedom?

It is often these dispositional traits that commentators have had in mind when they argue about moral enhancement. They are likely to have a biological basis, providing a target for engineering, and when they are present in people, it seems plausible that people are likelier than otherwise to act in ways that observers would understand as good. They can be thought of as patterns or structures of feeling and thinking that in various ways affect the value one attaches to things or people. People who might be

described as sympathetic, trusting, or generous can, of course, still make moral mistakes. As Rob Sparrow (2014) notes, it would be a remarkable drug 'that made us feel love only for what is worthy of love'. Arguably, though, people who are quite ready to love other people are likelier to behave well.

It is also the prospect of engineering the moral dispositions that arouses the fear that moral enhancement will undermine freedom and agency. In Harris' view, for example, changing these dispositions would predispose people towards acting in the right ways by altering their mindsets (by changing how they see and react to the world), when the goal should be to enhance their deliberative powers—their capacity to choose from among a range of possible actions. 'One thing we can say with confidence is that ethical expertise is not "being better at being good", rather it is being better at knowing the good and understanding the good. The space between knowing the good and doing the good is a region entirely inhabited by freedom' (Harris 2011, p. 104). Losing that space is fatal to morality. Harris, quoting Milton's line that God made man 'just and right/Sufficient to have stood, though free to fall', argues for the necessity of 'existential freedom', 'of choosing (and changing at will) our own path through life' (Harris 2011, p. 103). He presents this capacity as both conceptually necessary for any understanding of morality and central to human flourishing, insofar as it is a source of 'exhilaration and joy'. Along similar lines, Sparrow (2014) holds that people with biomedically enhanced moral dispositions will not be acting for the right reasons: outwardly they will appear to be acting morally, but they will not be acting morally in the sense that they are inwardly motivated by the right reasons. Instead of being explained by reasons, their behaviour will be explained by physical causes—by a pill that has usurped the existential freedom of a being whose actions should be explicable on grounds of reasons.

The dispute here may hang partly on different views about what freedom is—how much independence from causal influence is required for freedom. The idea of enhancing a disposition is the idea of making one sort of action more likely than another. As Savulescu, Douglas, and Persson argue (2014), there is still room for moral deliberation and for acting on reasons. The morally enhanced agent has been made more likely to deliberate and act in some ways than in others, but that fact is not at odds with the sense of freedom and agency that is necessary for morality. Indeed, they maintain that enhanced moral dispositions could foster freedom and agency because a better moral imagination could help the agent better understand the full significance of different courses of action, thereby letting the agent choose between them more knowledgeably.

In this view, the enhanced disposition acts as a kind of nudge—not determining behaviour but promoting some kinds of behaviour over others. In Harris' and Sparrow's objections, however, freedom appears to be understood as rather starkly opposed to any causal influence that would steer one's thinking. By promoting some behaviours, the enhancement has weighted the causal dice, and in order to protect the freedom that is necessary for morality, the dice must be unweighted:

physical causes should not merely not limit autonomy by ruling out some moral choices but should not even impinge on autonomy by fostering good moral choices.

There is always a physical story to tell about a 'reasoned' decision, of course. In fact, there are two different kinds of physical story. One is a full explanation of the decision. This is the sort of story that is possible, in principle, if the universe is fully determined and all reasoning can be explained in terms of the interactions of chemicals and electrical impulses. (It is not necessary to say that all reasoning is in fact nothing but the interaction of chemicals and electrical impulses, since a Kantian defender of freedom of the will might be able to allow that both stories can be told side by side, one giving a phenomenal account of the will and the other a noumenal account (Kaebnick 2006).) The other story is about the physical influences and constraints that produce the sort of deliberative capacities commonly associated with 'rational' beings and that, in any given decision-making context, influence and constrain individual agents. This is the sort of story that someone who takes a compatibilist stand on freedom of the will and determinism will offer in order to make sense of rationality and freedom. People choose freely when their decisions are made under various conditions commonly thought necessary for free choice—when they understand their options and are not unduly coerced when they select one of them, for example.

In effect, one who accepts moral enhancements may simply have a more expansive view of the range of physical constraints that are consistent with freedom. Yes, enhanced moral dispositions steer choice, but they fit within the compatibilist story about physical influences that bear on, but do not unduly bind, a moral agent's choice. A loving person is likelier not to engage in revenge killing, but the disposition to be loving need not be seen as having taken over their thinking, in much the same way that a slightly depressed person is still held to a standard of good moral deliberation. Indeed, one might hold that fellow-feeling, far from being a barely tolerated emotional intrusion on moral deliberation, fosters it: perhaps it encourages a person to ponder moral questions and deliberate about them more carefully.

If one still argues that such dispositional alterations are impinging on freedom, then the disagreement may be at a still more fundamental level. Someone who thinks of moral deliberation as purely a matter of reasoning—'strictly cognitive' capacities, understood as being free of the emotions—might regard alterations to an agent's moral disposition as introducing something fundamentally exogenous to morality. Thus the debate about moral enhancement may hang to some degree on different views about what morality is. Much of the opposition to moral enhancement has been built on broadly Kantian lines about moral deliberation—in particular, the requirement for an internalist account of reasons and for a very robust understanding of human freedom. For Kant, a moral decision that is based on the agent's emotions is by definition not autonomous. But if one takes a Humean view of moral decisions, seeing them as flowing from the emotions, then the prospect of altering the agent's moral dispositions may be less threatening to human freedom and morality precisely because those alterations *are* moral enhancements. They are changes that occur within our moral

faculties rather than changes to something outside our moral faculties—changes that by definition can only undermine morality.

16.4 Would Moral Enhancement Open the Door to Other Forms of Enhancement?

There is another, entirely different way in which moral enhancement might seem at odds with morality. One might expect strenuous opposition to moral enhancement from anyone who is opposed to enhancement in general—that is, from someone broadly concerned about the reach of human industry and engineering, and eager to champion an attitude of cherishing and preserving the natural world as it is (human bodies and brains included). In fact, for someone who takes this view, moral agency might seem to be a particularly important part of the world as it is, and a topic where the resistance to altering the given world would be strongest.

Of course, there are different ways in which the preservationist account might be developed, and perhaps some do lead to the construction of a long levee against all innovation, but a thoughtful view of the preservationist stance need not go that way (Kaebnick 2014). Here are a few starting observations about how a more thoughtful account might go—just enough to explain why moral enhancement need not be unacceptable to someone who is generally opposed to enhancement.

First, showing that a moral concern to preserve nature can be a legitimate part of a person's overall understanding of how they ought to live requires some reconsideration of what underwrites moral commitments in general. The utilitarian and Kantian perspectives of much normative ethical theory—perspectives that trace all moral commitments back to the intrinsic value of some aspect of human cognition, such as happiness or autonomy—provide no grounds for thinking that there might be value in preserving, to some extent, the world as we find it. Environmental ethicists have sometimes responded to the problem by trying to show that nature has value objectively: human valuers find value in nature because value is, in fact, in nature, accessible to those with the right cognitive apparatus. This requires some rather complex meta-ethics, however, which must be bolstered by a fair amount of clever argument. To be sure, some complex meta-ethics and clever argument is needed to show that *anything* has intrinsic value in that strong sense, happiness and autonomy included. Another approach, then, is to suggest that nothing has value objectively, that valuers do not find value so much as construct it, and that in constructing it, a foundational role is played by human sentiment—intersubjectively re-evaluated, refined, and channelled, but sentiment nonetheless. In this second approach, which builds on a broadly Humean account of morality, what our moral commitments trace back to is just our most fundamental emotional commitments.

This way of understanding moral values seems to make room for a moral concern for nature. It also provides a target for moral enhancement, and it dodges the objection

that the emotions are usurping the place of reason in moral deliberation. If moral values rest on tested and developed sentiment, then deliberation about any given moral problem can and should involve emotion (as well as reason). Having good emotional dispositions would be critical to making good moral judgements, and developing good emotional dispositions would be critical to having them. Traditionally, the process of developing them depends on parenting, enculturation, and sometimes therapy, but, in principle, the traditional methods might be augmented with biotechnological interventions.

A second general observation about a thoughtful way of defending a preservationist stance towards nature is that the value attached to nature would be but one of a range of values. And if that value is explained along Humean lines rather than being underwritten by some very special metaphysics in which nature is seen as a special, objective source of value, then nature need not be sacred or inviolable. The desire to leave some natural places alone would compete, in a rough way, with the need to ensure human well-being, to distribute goods justly, and so on—and so sometimes natural places could appropriately be 'improved' and natural states of affairs could appropriately be 'enhanced'. The countervailing moral considerations may be greater. (Indeed, far from taking the role of moral trump, concerns about nature are likelier to take a back seat to other concerns, and perhaps the champion of preservationism would even come up short of regarding their preservationist stance as flatly obligatory for all other people [Kaebnick 2014, pp. 46–61].) If human nature leaves much to be desired in terms of humaneness, then maybe the prospect of moral enhancement is one of those instances in which the countervailing moral considerations are urgent.

Nor need we suppose that valuing nature is the same thing as thinking that natural states of affairs always work out just right for human well-being. Human suffering seems to be a part of the world, and in principle, confronted with a problem in which an intervention into nature would make possible a reduction of human suffering, one might reasonably decide to intervene. That's what medicine is. Those who are concerned with preserving natural states of affairs need not reject all uses of medicine; they only ask questions about how far we want to go with medical interventions. It's one thing to try to prevent or interrupt diseases that would knock one off before one reaches that stage of life traditionally known as 'old age', and another to prevent or interrupt the ageing process so that one can live well past that age (and without ever seeming to enter it). It may be necessary not to die during one's young adulthood in order to have a good life, but it has generally not been thought necessary to live past one's old age to do so. But, arguably, moral enhancement is becoming necessary in order for people to live together on one planet and not to ruin it.

Third, how we judge an appeal to preserve a natural state of affairs may also vary by context. Concerns about altering nature arise in various domains: from human enhancement of various kinds and in various contexts, through human manipulation of non-human organisms in agriculture and industry, and on to the human relationship to the environment. Although it would be a merit of any philosophical account of

moral concerns about altering nature that it rested on a theory of values and a concept of nature that provided a way of critically thinking about all of those disparate concerns, we need not come to the same conclusions about the merits of those appeals. As Thomas Murray (2007) has argued, one should consider the point of the overall practice in which a given human activity is a candidate for enhancement. One might take a very different view about the desirability of human enhancement in a surgical theatre compared with a sporting arena: one hopes for the best results possible in surgery, but one might enjoy sports precisely because of the way it features familiar, ordinary human bodies.

Moral enhancement therefore needs to be considered on its terms, not just as another instance of human alteration of the natural world, and (as Harris shows) it needs to be considered in light of how it supports or conflicts with the relevant human practice. So what is the point of morality? On the one hand, we would certainly not want to create better human beings if they no longer counted as human beings at all. On the other hand, there might, in principle, be some moral enhancements that would not be dehumanizing, and we really do very much want to be able to live together on this one planet.

These three observations about one possible account of the preservationist stance towards nature push in two, partially contrasting, directions: they are part of an argument for making sense of the preservationist stance, but they also limit the force of that stance. The first observation, about the nature of values, puts a moral concern about nature on the same conceptual footing as other values, but the second and third suggest that it is limited by, and balanced against, other values. There is perhaps a sense in which a moral concern about preserving nature ends up a bit downgraded in this account. It is not grounded in some deeper or higher, morally prior truth, as might be the case if the argument were that humans should not alter nature because nature is the ultimate source of moral rules or values, an eternally fixed ground of moral order, a reflection of God's plan, or some such. But that lofty picture has never looked very plausible, and such a strong view about the value of nature would require a very strong defence. Thus downgrading a concern to preserve nature does not, ultimately, push against the argument for making sense of it. It is in part precisely by making it a plausible, ordinary part of the moral life—not downgrading it, let us say, but bringing it down to earth—that we make sense of it. (So understood, the commitment to nature reflects the loneliness of the human condition: it is grounded in the meaning people attach to nature in a universe that, considered in itself, is devoid of meaning, and perhaps it is partly through that expression of acceptance and solidarity with nature that people redress their loneliness.)

The argument so far is that the preservationist stance is logically consistent with altering the moral dispositions. But there is another, circuitous way in which altering our moral dispositions might be at odds with a commitment to accept and preserve the world as it is. Altering a person's moral dispositions might lead to changes in their understanding of meaning, in their self-understanding, and in their sense of the human

relationship to nature. And if moral values are socially constructed, then maybe they could be *reconstructed* in substantially different forms. In short, a commitment to preserve the world is itself embedded in our existing moral dispositions, and therefore altering our moral dispositions might run the risk of undermining, even choking off, a commitment to preservation.

The structure of this worry takes the form of a slippery slope: if we accept so and so, then we might eventually find ourselves accepting thus and such. The thought is not that accepting some version of moral enhancement logically commits one to abandoning all commitments to the preservation of nature, but that one thing could lead to another, there might not be footholds along the way, and we might end up doing so anyway. But there are in fact footholds. They are simply the moral values that prompt us to think about moral enhancement in the first place. The goal of moral enhancement would be to better live up to those values, to honour them more closely; and while our standards for living up to them 'better' might change, the underlying moral values themselves need not. Those values would be the framework for moral enhancement in the first place. In fact, revising our moral judgements would be likelier if values rested on some objective foundation: a morally enhanced person would perceive and reason about that foundation more clearly, and might learn that some long-accepted moral views are simply wrong.

Strictly speaking, the underlying moral values need not change if the moral enhancement is carried out correctly. A version of the slippery slope worry survives this point: perhaps moral enhancement would lead off in new directions, either because our values were not really guiding the process or because the process was not adequately understood and an alteration that had been thought to have one effect on the moral dispositions had other effects as well (or instead). There is not a single biological mechanism or a single gene involved in morality, and the changes we could make—if we could make changes at all—would come about in bits and pieces, and fits and starts. We would not be enhancing morality simpliciter, but this or that specific characteristic that we think is part of moral behaviour. And we might well misunderstand the changes, learning only after the fact exactly what effect they had. That may, in fact, be the likeliest course of events if we start altering moral behaviour. However, this sort of problem is not a special problem with moral enhancement. It does not, for example, involve the sort of internal inconsistency that Harris argues afflicts proposals for moral enhancement. It is a version of a fairly familiar problem about the governance of technology development. It points up the need to go slowly and carefully.

16.5 What Would it Mean to Do Moral Enhancement Well?

Much of the discussion so far has turned on some ambiguity about what moral enhancement would be enhancement of. I have argued that moral enhancement would not necessarily undercut a broad-but-not-across-the-board opposition to enhancement

because it would be enhancement of our capacity to act on our existing values, and some modest ideal of preserving nature might be one of those values. I have also argued that the debate about whether moral enhancement is at odds with freedom may turn on a broad dispute about what moral deliberation is—about whether moral deliberation should be reason freed of sentiment or sentiment conjoined with reason.

Pointing out that broad divide only begins to get at the problem of identifying moral enhancement, though. There is an ever-growing literature, emanating from moral psychology and neuroscience, and much discussed in behavioural economics and in the literature on risk perception, on various patterns of thinking that affect moral decision-making. The core theme in much of this work is that moral decision-making typically relies on the emotions as well as on reason, and that reason often has relatively little to do with it. As noted in Section 16.2, Joshua Greene suggests that moral decision-making can be done in a slow, deliberative mode or a fast, more intuitive and emotional mode. In Jonathan Haidt's (2013) metaphor, moral judgements flow from both reason and emotion, but reason has about as much control over emotion as a person has over an elephant they are riding: the rider can nudge the elephant, but the elephant usually goes where it wants.

The attraction, or threat, of moral enhancement is the possibility of changing these metaphors. But what metaphor should we want? How much control should the sentiments have, and which of the judgements that they tend to promote should be preserved or fostered? If we have a choice, should we enable the rider to take more control of the elephant, and where do we want it to go?

Greene is clear that reason is superior to the emotions and should have a more prominent role. The only invaluable contribution of the emotions is to provide us with the ultimate moral goal of wanting happiness. It is the job of reason to achieve that. Haidt tells a more complicated story about what sentiment provides, and he is attracted to the richness and diversity he finds. He describes rationalist morality, with its 'unusually narrow' focus on autonomy and harm, as Western, educated, industrialized, rich, and democratic (WEIRD; 2013, p. 96). For most cultures and many people even in Western nations, there is an assortment of basic moral concerns pertaining to caring, liberty, fairness, loyalty, respect for authority, and sanctity. These six concerns, rooted in moral sentiments, include but go beyond the concerns of WEIRD morality.

The question for moral enhancement, then, is whether WEIRD morality is a refined and purified morality that we should extol and extend, or a degraded, desiccated form that should be enriched and enlivened. The answer is not clear. Haidt allows only that for implementing public policies in Western democratic societies 'there is no compelling alternative to utilitarianism. [. . .] [L]aws and public policies should aim, as a first approximation, to produce the greatest total good', as long as the public understanding of the greatest good is informed by a recognition that, within a morally diverse society, the citizens will understand the good in ways that connect to various sentiments (2013, p. 272).

Another way of thinking about whether to redirect the elephant is to consider the 'biases' and preferences that are the subject of behavioural economics and the psychology of risk perception. In the risk perception literature, for example, Paul Slovic (1987) has argued that whether risks are perceived as greater or lesser does not track a potential harm's severity (which can be measured in various ways but is often pegged to the number of deaths caused) and likelihood. Instead, risk perception is also shaped by considerations such as the degree to which a potential harm is familiar or unfamiliar (e.g. whether a cancer risk is traced to grilled food or to radiation), and the extent to which it seems to be under the control of the person harmed or beyond that person's control (think here of how death in an automobile accident might compare with death in a plane crash). Behavioural economists have shown that, when making quick decisions about costs and benefits, people tend to reveal an assortment of biases, such as 'status quo bias' (a preference for the status quo and a resistance to change) and 'loss aversion' (a tendency to assign more value to something one already possesses than to something one has the prospect of acquiring) (Kahneman, Knetsch, and Thaler 1991).

Some subjective preferences look indefensible and would seem to be good targets for moral enhancement. Douglas proposes the elimination of racial bias as a reasonable early goal for moral enhancement (2011, p. 470). Another might be 'availability bias'—the tendency to assign more weight to something which is fresh and vivid in one's mind because one recently experienced it or knows more about it. The fact that a given risk or benefit can be thought of more clearly does not seem to be a good reason for thinking it is a *greater* risk or benefit.

On the other hand, consider the bias for 'identified' over 'statistical' lives. It is a matter of great perplexity to rationalist thinkers that people will often go to greater lengths and spend more money to save identified victims than they will to save a comparable or greater number of people who can be only numbered, not named. We will do more to rescue trapped miners, for example, than to prevent the mine accidents that are sure, over time, to entrap miners. 'Objectively' it cannot make sense to assign more value to people whose names and faces are known than to those whose names and faces are unknown. The economist Charles Fried reportedly held that we should even be willing to divert money from the trapped miners to efforts to prevent future mine accidents, and also that we should be willing to tell the trapped miners that that's what we intend to do (Slote 2015). But to one who thinks that moral decisions depend heavily on subjective, sentiment-based reactions, such as empathy, matters will look different. The 'virtue-ethical sentimentalist' Michael Slote has argued that we are more eager to save trapped miners than to avert future mining accidents because 'empathy is more sensitive to what is contemporaneous and definite than to what is future and indefinite' (2015, p. 151).

Slote frames the contrast between his and Fried's views as a kind of causal point: Fried just doesn't understand where values come from, but once one is clear on that, the eagerness to save the known miners makes sense. However, as Slote well knows, the interesting moral question is ultimately about which view is justified, and the

challenge for moral enhancement is to think through such justificatory questions. A moral enhancement programme could have to decide whether to adjust morality in the direction of Fried. So, is our goal strictly to do those things that bring about the greatest state of happiness in the world, or do we care also about the human capacity for empathy? And if the emotions are the basis for morality, then how could we *not* care about human empathy?

Similar things could be said about many other basic moral sentiments. Here, then, is a problem to layer on top of the problem I noted at the beginning. To engage in moral enhancement, we would have to know how to do it. I started by noting that, technically, we do not yet know how to do it and might be a long way from knowing how to do it. But we also do not know how to do it conceptually. The disagreement about whether to engage in moral enhancement has often been read as a disagreement about the merits of enhancement. In fact, I have argued, it may turn in good part on different views about the merits of different aspects of moral decision-making—the relative roles of reason and emotion, and the relative roles of the different moral values that a sentimentalist is likely to see as playing a vital role in moral thinking. The technical difficulty means that a programme of moral enhancement could easily be done very poorly and would require good oversight. The implication of the conceptual problem for moral enhancement is that doing it in a way that was universally accepted *as* done well might be more or less impossible, not just because moral debate is hard but because, if moral deliberation has a lot to do with the emotions, morality may be too rich and diverse—too complicated—to see how to undertake a clear social programme of moral enhancement.

References

Agar, Nicholas (2007). 'Whereto Transhumanism? The Literature Reaches a Critical Mass', *Hastings Center Report* 37(4): 12–17.

Agar, Nicholas (2014). *Truly Human Enhancement: a Philosophical Defense of Limits*, Cambridge, MA: MIT Press.

Douglas, Thomas (2011). 'Moral Enhancement', in Julian Savulescu, Ruud ter Meulen, and Guy Kahane (eds), *Enhancing Human Capacities*. Malden, MA: Wiley-Blackwell, pp. 467–85.

Greene, Joshua (2013). *Moral Tribes: Emotion, Reason, and the Gap between Us and Them*. New York: Penguin Press.

Haidt, Jonathan (2013). *The Righteous Mind: Why Good People Are Divided by Religion and Politics*. New York: Knopf Doubleday.

Harris, John (2011). 'Moral Enhancement and Freedom', *Bioethics* 25(2): 102–11.

Kaebnick, Gregory E. (2006). 'Behavioral Genetics and Moral Responsibility', in Erik Parens, Audrey R. Chapman, and Nancy Press (eds), *Wrestling with Behavioral Genetics: Science, Ethics, and Public Conversation*. Baltimore, MD: Johns Hopkins University Press, pp. 220–34.

Kaebnick, Gregory E. (2014). *Humans in Nature: The World As We Find It and the World As We Create It*. New York: Oxford University Press.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). 'Anomalies: the Endowment Effect, Loss Aversion, and Status Quo Bias', *Journal of Economic Perspectives* 5(1): 193–206.

Murray, Thomas H. (2007). 'Enhancement', in Bonnie Steinbock (ed.), *The Oxford Handbook of Bioethics*. New York: Oxford University Press, pp. 491–515.

Parens, Erik (2005). 'Authenticity and Ambivalence: Toward Understanding the Enhancement Debate', *Hastings Center Report* 35(3): 34–41.

Parens, Erik (2009). 'Toward a More Fruitful Debate about Enhancement', in Julian Savulescu and Nick Bostrom (eds), *Human Enhancement*. Oxford: Oxford University Press, pp. 181–98.

Parens, Erik (2015). *Shaping Our Selves: On Technology, Flourishing, and a Habit of Thinking*. New York: Oxford University Press.

Persson, I. and Savulescu, J. (2012). *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press.

Savulescu, J., Douglas, T., and Persson, I. (2014). 'Autonomy and the Ethics of Biological Behavior Modification', in Akira Akabayashi (ed.), *The Future of Bioethics: International Dialogues*. Oxford: Oxford University Press, pp. 91–112.

Slote, Michael (2015). 'Why Not Empathy?', in I. Glenn Cohen, Norman Daniels, and Nir Eyal (eds), *Identified versus Statistical Lives: An Interdisciplinary Perspective*. New York: Oxford University Press, pp. 150–60.

Slovic, Paul (1987). 'Perception of Risk', *Science* 236: 280–5.

Sparrow, Rob (2014). '(Im)Moral Technology? Thought Experiments and the Future of "Mind Control"', in Akira Akabayashi (ed.), *The Future of Bioethics: International Dialogues*. Oxford: Oxford University Press, pp. 113–19.

17

The Evolution of Moral Enhancement

Russell Powell and Allen Buchanan

17.1 Introduction: Moral Enhancement, Evolution, and the Liberal Project

Human beings in the twenty-first century are confronted with a daunting array of moral problems, from climate change, poverty, and genocide to the prospects of nuclear war and terrorism—ethical challenges that human moral psychology, which evolved to function under very different social and technological circumstances, is arguably ill-equipped to address. In light of this evolutionary psychological mismatch and the seriousness of the threats we now face, some moral philosophers have looked to the use of biomedical technologies—such as neurological, pharmacological, and genetic interventions—to enhance human moral capacities, including moral emotions and morally relevant cognitive abilities (cf. Persson and Savulescu 2008, 2012, 2014; Harris 2010; Buchanan 2011; Agar 2013). Much of this philosophical attention to biomedical moral enhancement (BME) has been motivated in part by the belief that cultural forms of moral enhancement (e.g. moral education) have been only moderately effective and are simply not up to the task of mitigating major anthropogenic harms and existential risks (see e.g. Persson and Savulescu 2008).

A key framing assumption underlying the BME project so conceived is that evolved human moral nature is a source of great, if not insurmountable, resistance to solving the onerous moral tasks that lie before us. This has led some liberal moral philosophers to conclude that the prospect of major moral gains and the avoidance of major moral losses are unlikely if we do not fundamentally address the discrepancy between our prehistoric moral psychology and the ecologically alien world that human beings now inhabit. One way of aligning human moral psychology and human moral ecology would be to radically alter our social and technological environment so as to return to pre-industrial—or even pre-agricultural—human lifeways. This option is neither plausible nor on most accounts desirable, and thus, seeing no other alternative, some authors have considered the possibility of altering human biology to meet the ethical demands of the modern world.

The central question we wish to explore here is this: how much moral progress is possible, and can major moral catastrophes be averted, without the biomedical enhancement of human moral capacities? One way of approaching this question is to look at the extent of moral progress that has already been achieved, as this may give us some idea as to the power and limits of cultural moral enhancement. It is only by connecting up our history of moral achievements with empirically rigorous investigations of human morality, however, that we can begin to make meaningful projections about the ultimate scope of moral progress with, or without, BME.

In Section 17.2 we summarize the prevailing evolutionary account of morality, and explain why it is unable to accommodate cases of sweeping, progressive moral change that we refer to collectively as the 'inclusivist anomaly'. We show that both conservative and liberal authors, whom we refer to as 'evoconservatives' and 'evoliberals', respectively, have relied on an evolutionary account of human moral nature that is called into question by this inclusivist shift. We then go on to sketch an evolutionary account of moral psychological development that not only is consistent with the prevailing evolutionary explanation of morality but also accommodates the shift towards inclusivity. In particular, we propose that 'exclusivist' (parochial, group-based) morality is the result of an adaptively plastic (conditionally expressed) moral response that is sensitive to environmental cues that were historically indicative of out-group threat. Unpacking this complex norm of reaction is crucial, we believe, to understanding the evolved constraints on moral progress and how they have been, and may be further, overcome.

In Section 17.3 we go on to consider the implications of this evolutionary model of moral development for the relative efficacy and plausibility of BME as a solution to some of the most pressing moral problems of our time. Once these problems are recast in terms of moral inclusivity, it becomes clear that BME technology, as narrowly conceived by BME proponents, is unlikely to be necessary or effective when it comes to addressing them. According to the model we propose, efforts to achieve major moral progress and avert serious moral catastrophes do not go against the human evolutionary grain *tout court*; they only go against the evolutionary grain under certain conditions, and these conditions are both epistemically accessible and within our practical powers to modify. We conclude that cultural innovations that make use of our best understanding of the evolutionary development of human morality stand the best chance of meeting the above ethical goals.

17.2 Towards an Evolutionary Developmental Model of Moral Progress

17.2.1 The Received Selectionist Explanation of Morality

Why did morality originate and what functional purpose might it have served?[1] The prevailing evolutionary explanation of morality holds, in brief, that morality evolved

[1] Arguments in Section 17.2 draw upon ideas presented in Buchanan and Powell (2015) and Buchanan and Powell (forthcoming).

via natural selection in small to moderate-sized hunter-gatherer groups because it promoted higher levels of within-group cooperation. It did this by reducing free-riding and mitigating selfish tendencies, which it accomplished both internally through individual moral emotions and externally through institutionalized incentives that skewed the pay-off matrix towards cooperation. This increase in cooperation gave groups with effective moralities several fitness advantages over non-moral and less moral ones. These included greater foraging yields, enhanced warfare and territorial acquisition capabilities, better resource management, and more efficient dispute resolution. These beneficial properties in turn permitted larger group sizes, which would have fed back into the aforementioned advantages, and so on.

If the dominant evolutionary account is right, then morality has a dark side. Modelling, ethnographic, archaeological, and psychological data suggest that morality evolved in an arena of competing cultural groups characterized by frequent and often lethal intergroup conflict. Groups that contained more altruists and more individuals willing to punish moral norm infractions, so the theory goes, tended to replace groups with less effective moralities in economic and military contests between groups (Richerson and Boyd 2005).

So a key implication of the prevailing evolutionary explanation is that for altruism and other pro-social tendencies to have evolved beyond kin, they had to be selectively restricted to members of a cooperative group. Just as free-riding undermines group performance, so too does extending altruism to competing groups that might otherwise be exploited, marginalized, or eliminated. As human group size increased due to the cooperative functions of morality, linguistic and ethnosymbolic markings became increasingly important for drawing group boundaries and demarcating the moral community.

Thus, according to the received evolutionary account, morality originated and evolved in an environment in which out-groups represented significant threats. One such threat concerned economic and physical competition between groups over scarce and scattered resources—a condition that appears to have driven the evolution of intergroup raiding and even extreme violence in common chimpanzee populations, while an ecology of abundant static resources drove the evolution of comparably peaceful social relations in bonobo chimps (Wrangham and Peterson 1996). A second out-group threat was the risk of social parasitism: whereas effective institutions exist within egalitarian hunter-gatherer groups to prevent exploitation (Boehm 2001), the lack of institutions at the metagroup level to ensure mutually beneficial cooperation among groups created a risk of free-riding that made intergroup cooperation hazardous. A third major out-group threat was the risk of biological parasite transmission between groups that lacked immunity to foreign infectious agents. The risks of social and biological parasitism provided further ecological conditions conducive to the evolution of intergroup antipathy, aggression, and distrust.

17.2.2 The Evoconservative–Evoliberal Convergence

If the received evolutionary account of morality is right, then it would suggest that exclusivist moral tendencies dominated the vast majority of human moral history.

Furthermore, it would appear to bode poorly for one important dimension of putative moral progress—namely, the possibility of moving towards more inclusivist moralities.[2] By 'inclusivist' moralities we mean those that reject group-based restrictions on membership in the moral community, such as those based on race, ethnicity, gender, species, or the self-serving cooperative relationships between groups, and that avoid arbitrary discounting of the interests of the members of out-groups.

Given this reading of human evolutionary history, it is perhaps not surprising that some scholars, whom we have elsewhere labeled 'evoconservatives' (Buchanan and Powell 2015), have attempted to draw political philosophical lessons from the received evolutionary explanation of morality. Evoconservatives claim that because human pro-social tendencies are restricted, *pace* evolutionary theory, to one's own kin or identifying group, certain types of moral progress are inappropriate or unrealistic for beings like us. In particular, they argue that there are good evolutionary reasons to think that inclusivist moral principles, which extend equal moral regard beyond the confines of the group, are psychologically implausible, at least in the absence of reciprocal benefit (Asma 2012, pp. 44–6; Haidt 2012, p. 245). In this way, evolutionary theory lends scientific credence to longstanding but historically underevidenced conservative suspicions about the limits of human altruism and the frailty of non-strategic moral relations between peoples.

The notion that there are strong evolutionary constraints on inclusivist morality is not unique to the conservative tradition. Some liberal philosophers, whom we shall call 'evoliberals', similarly argue that the moral legacies of human evolutionary history make it difficult to act on liberal moral norms insofar as they are inclusive. Contra evoconservatives, however, evoliberals contend that rather than giving us reason to pare back our norms, these evolutionary constraints suggest that a systematic programme of BME will be crucial in order to drive major moral progress and avert future moral catastrophes. Evoliberals thus conclude that BME should not only be pursued but also be afforded a significantly higher priority than it is currently given in the allocation of limited social resources (Persson and Savulescu 2014). In a passage worth quoting at length, Ingmar Persson and Julian Savulescu (2012) sum up this line of argument, which they canvass over a series of joint papers:

> For most of the time the human species has existed, human beings have lived in comparatively small and close-knit societies, with primitive technology that enabled them to affect only their most immediate environment. Their moral psychology adapted to make them fit to live in these conditions. This moral psychology is 'myopic', restricted to concern about people in the neighbourhood and the immediate future. But through science and technology, humans have radically changed their living conditions, while their moral psychology has remained fundamentally the same throughout this technological and social evolution, which continues at an accelerating speed. Human beings now live in societies with millions of citizens and with an advanced scientific technology which enables them to exercise an influence that extends all

[2] For a discussion of different types of moral progress, see Buchanan and Powell (unpublished manuscript).

over the world and far into the future. This is leading to increasing environmental degradation and to harmful climate change. The advanced scientific technology has also equipped human beings with nuclear and biological weapons of mass destruction which might be used by states in wars over dwindling natural resources or by terrorists. Liberal democracies cannot overcome these problems by developing novel technology. What is needed is an enhancement of the moral dispositions of their citizens, an extension of their moral concern beyond a small circle of personal acquaintances, including those existing further in the future. The expansion of our powers of action as the result of technological progress must be balanced by a moral enhancement on our part. Otherwise, our civilization, we argued, is itself at risk. It is doubtful whether this moral enhancement could be accomplished by means of traditional moral education. There is therefore ample reason to explore the prospects of moral enhancement by biomedical means. (pp. 399–400)

On the evoliberal view, given the fixed (or highly recalcitrant) nature of human moral psychology, the most effective means of transcending our inability to extend moral concern beyond the group, including to individuals of future generations, is not sociocultural but biological, in the form of BME. Evoliberals thus presuppose a degree of 'moral hardwiring' that is significantly determinative of the shape of human morality, robust against changes in the moral developmental environment, largely insensitive to rational persuasion, and a perilously poor fit for the moral challenges that face humanity in the new millennium. Evoconservatives join them, at least on the first three points. In the evoconservative/evoliberal view, inclusivist morality is simply a 'façade' covering an essentially xenophobic nature (Persson and Savulescu 2014, p. 105).

We might distinguish a weaker evoliberal claim, which holds simply that our moral problems are so momentous that any possible means of reducing our parochial tendencies should be on the table, BME included. We are happy to embrace this weaker evoliberal claim, but note that it is not a very interesting argument since it does not make any substantive assertions about how important BME is likely to be given our evolutionary history, and thus it does not say anything about how BME should be prioritized relative to 'traditional' modes of moral enhancement. BME carried out on a global scale comes with significant risks of unintended consequences, as well as the harms associated with incentivizing its use at population levels; and thus the advocates of BME must envision its prospective pay-off as sufficiently great, and the expected utility of traditional cultural moral enhancement as sufficiently minimal, to justify taking these risks. Indeed, the assumption of cultural impudence is crucial, because even if it turns out that we have weighty moral reasons not to engage in BME—or if BME turns out to have significant ineliminable risks—these reasons and risks will be outweighed by the need to use BME to avoid our greatest existential threats.

17.2.3 Reasons to Doubt the Strong Evolutionary Moral Constraints View

If the evoconservative/evoliberal line is right, then human evolutionary history exerts significant constraints on the way humans are disposed to delineate the moral community. Moreover, it might be read to suggest that the only type of morality that humans are

meaningfully capable of engaging in is a strategic morality grounded in cooperative group reciprocity. In this type of morality, moral consideration is extended beyond close kin only to individuals who are capable of either contributing to or disrupting cooperation. It is no accident that evolutionary theories of moral psychology, as well as game theoretical approaches to the evolution of morality, have tended to operate with such a strategic, self-interested conception of human morality (e.g. Alexander 1987; DeScioli and Kurzban 2013). It also explains why some eminent evolutionary theorists have been famously sceptical of the existence of genuinely altruistic moral motivations (see e.g. Ghiselin 1974, p. 247).

This strategic, Hobbesian conception of morality represents an impoverished view of what human morality now encompasses. Contemporary human morality is far more inclusive and non-strategic than evolutionary theories of morality would predict. This in turn suggests that the 'strong evolutionary constraints' view of human moral nature is mistaken, and it leads us to propose a very different way of thinking about the nature of human moral development. We will first highlight several evolutionary moral anomalies that call into question the strong evolutionary constraints view. We will then go on to sketch an evolutionary model of moral development that can accommodate these anomalies, and then bring this model to bear on ethical questions surrounding BME.

One feature of contemporary human morality that stands out as anomalous from the standpoint of the strong evolutionary constraints view is that non-human animals are now widely regarded as proper subjects of moral worth. The prevalence and strength of this inclusivist belief is reflected in domestic and supranational legal regimes governing animal cruelty, animal welfare, animal experimentation, and animal exhibition. Protections for animals have been enacted and enforced at considerable expense, particularly in the agricultural and experimental contexts, notwithstanding the opposition of special interest groups. Non-human animals possess no strategic capacities and have been readily exploited by humans in fitness-enhancing ways for thousands of years, and thus these more inclusive norms have arisen despite evolutionary constraints on moral inclusivity. It is true that non-human animals are still treated as mere units of production and their interests are largely disregarded in substantial segments of the world. But these facts do not belie this important inclusivist trend any more than the fact that many countries still deny basic civil and political rights undercuts the global trend in the expansion of basic liberties.

The non-human animal case is particularly important because it shows that some inclusivist shifts are not simply the readjustment of groupish morality to accommodate a larger group of cooperators (see e.g. Gaertner and Dovidio 2000). The ethical treatment of non-human animals is motivated in part by the recognition that many of them have cognitive properties (e.g. sentience) that deserve moral consideration, even when they have no cooperative relationship and indeed little else subjectively in common with human beings.

Second, significant numbers of people now regard many moral norms as applicable to all people, rather than restricted to members of a given race, gender, religion, ethnicity, or nationality, as might be expected if evolutionary forces strongly constrained human moral response and the delineation of the moral community. Some of these universalized moral norms fall under what we take to be a third evolutionary moral anomaly—namely, the culture of human rights. It is now widely held that all human beings are entitled to be treated in ways that respect their basic interests regardless of which ethnic, racial, or religious groups they happen to belong to, and regardless of their age or able-bodiedness. There is now robust domestic and international legal support for this inclusive moral commitment, which has been shown to significantly shape the behaviour of states (Goodman and Jinks 2013). Again, this is not to say that these changes are universal, but the shift towards more inclusive moral norms is not merely academic or inspirational. The scope of this shift and its institutionalization in law and policy is sufficiently large to view these changes as progressive developments. Common to each of these inclusivist shifts is that they are premised on a 'subject-centred' or non-strategic conception of basic moral status, which compels us to recognize the moral standing of individuals even when they pose no threat to us or do not contribute to cooperative goods (Buchanan 1990).

We think that none of these inclusivist trends, including the broader shift towards a subject–centred morality that underpins them, can be explained as adaptations, or even constrained byproducts of adaptations, either to ancestral or more recent environments (for an extended discussion, see Buchanan and Powell 2015). This suggests to us that evolved constraints on moral inclusivity are far weaker than evoliberals or evoconservatives suppose. What then can account for the shift towards moral inclusivity?

It is tempting to conclude that inclusivist moral progress since the Enlightenment can be attributed to the responsiveness of human moral thought and behaviour to reasons (Singer 2011; Lazari-Radek and Singer 2012), to the capacity for moral self-scrutiny (Moody-Adams 1999), or to the ability to recognize and resolve moral inconsistencies (Campbell and Kumar 2012). We agree that reason, and, in particular, what we have called 'open-ended normativity'—the ability to subject one's norms and even one's conception of who has moral standing to rational scrutiny, conclude they are arbitrarily restrictive, and revise them accordingly—is an important part of any explanation of inclusivist moral progress. However, we agree with evoliberals that reason alone is not a sufficient platform from which to address the major moral challenges we face (e.g. Persson and Savulescu 2014, p. 109). This is not because reason plays a largely epiphenomenal role in moral judgement, as some empirical moral psychologists have claimed (e.g. Haidt 2001). Rather, explanations that appeal solely to rationality fail to account for inclusivist moral progress because they are crucially incomplete: humans have possessed rationality for hundreds of thousands of years, and yet the inclusivist shift has taken place only in the most recent eye blink of human evolutionary history. Furthermore, reason often acts to justify pre-existing exclusivist moral commitments and can even drive moral regressions: consider, for instance, 'rational' appeals to natural

differences between groups that are thought to justify excluding certain people from the moral community (e.g. beliefs that certain groups have inherently attenuated capacities for rationality, intelligence, or the ability to feel pain). Thus if reason does play a key role in driving inclusivist moral progress, it only does so under certain, as yet unspecified, conditions. Our aim in Section 17.2.4 is to flesh out these conditions by proposing an evolutionary model of moral psychological development that can accommodate inclusivist moral developments.

17.2.4 A Dynamic Evolutionary Model of Moral Psychological Development

Some species of water flea develop elaborate armour, including defensive spines and helmets, only if they detect the chemical correlates of a predator in the water in which they grow. The development of armour, however, comes at a cost, such as reduced locomotion and added energy requirements. As a result, water fleas have evolved a conditionally expressed trait that enables them to express armour only when they find themselves confronted with the high probability of a predator-rich environment. Such traits are known in evolutionary biology as 'adaptively plastic' traits. Moreover, such environment-induced epigenetic changes can be inherited across several generations; in the case of water fleas, the epigenetic inheritance of armor results in the trait being exhibited by several subsequent generations of water fleas even though they were not exposed to the triggering chemical.

An examination of a range of evidence from history, psychology, biology, anthropology, and cognitive science leads us to propose that exclusivist moral psychology is an adaptively plastic trait, like flea armour, where the relevant threat relates to out-group predation and competition. We make this case more fully elsewhere (Buchanan and Powell forthcoming), and for reasons of space provide only a sketch of the argument here. Our idea is that exclusivist moral response is a conditionally expressed trait that develops only when cues that were in the past reliably correlated with out-group threat are detected. As with flea armour, the development of exclusivist moral tendencies has costs. In particular, out-group aggression, antipathy, and distrust reduce the chances of mutually beneficial interactions with neighbouring groups, such as trade and alliances. Because of this evolutionary trade-off, exclusivist tendencies will, so the theory goes, be tempered in environments in which out-group threats are not detected during development.

In the animal world, the adaptively plastic detection of a predation threat can involve not only the detection of pertinent chemical cues but also the inspection of predatory types, motivations, and behaviours. Because humans are linguistic and robustly cultural, the detection of out-group threat can also involve the social transmission of beliefs about out-groups. This will be an important feature of our evolutionary model of moral development, and a central reason why BME is likely to be of limited efficacy in averting moral tragedies.

Evolutionary psychologists Corey Fincher and Randy Thornhill (2011) propose an adaptive plasticity hypothesis to explain the strong cross-cultural correlation between

what they call 'in-group assortative sociality' and parasite stress. They marshal a formidable amount of evidence in support of the claim that in-group bias tends to develop when signs of infectious disease are detected during human development, whereas less xenophobic attitudes and behaviours (or, in our terminology, more inclusivist ones) tend to emerge when cues of infectious disease are absent. We suggest expanding this account to include other signs of out-group threat broadly construed, such as competition for scarce resources and, especially, socially constructed beliefs about out-groups. In so doing, we carve out a more fundamental role for culture in an adaptive plasticity account of moral psychology.

We will defend two hypotheses. The first is that inclusivist morality is a luxury good. By this we mean that developmental environments that are favourable to inclusivist morality are rare, having emerged only recently in human history and remaining largely confined to well-resourced populations with robust healthcare infrastructures, markets, rule of law, reduced rates of criminality, high rates of education and literacy, and so forth. Indeed, if the adaptive plasticity account of moral psychological development is right, then inclusivist moral progress and the above sociopolitical developments do not simply have a common cause. They are reciprocal causes of one another: conditions of infectious disease, resource scarcity, physical insecurity, interethnic conflict, and low rates of productivity seed exclusivist moral responses, which in turn exacerbate the conditions that act as triggers for the development of exclusivist tendencies and their attendant psychological orientations, such as mental rigidity, closed-mindedness, and dogmatism. Under these conditions, subjecting moral norms to rational scrutiny and institutionalizing more inclusive norms will be particularly difficult.

By the same token, cultural innovations that alleviate the above trigger conditions act to break this vicious spiral, creating an environment in which reasoning about moral norms, and thus inclusivity, can flourish. Cultural innovations can do this in two ways. First, they can ameliorate the harsh conditions of the environment in which exclusivist morality flourishes, such as through improvements in productivity, the creation of institutions that reduce intergroup conflict, and advances in medicine and public health infrastructure that reduce parasite stress. Second, once these inclusivist-friendly conditions have been established, the introduction of new moral norms can begin to shift the moral psychology of human populations in more inclusivist directions. Whereas Persson and Savulescu view normative constructs such as the Golden Rule as reflections of our hardwired moral nature (2008, p. 168, 2014, p. 22), we see them as progressive, culturally scaffolded innovations which, under the right conditions, have the potential to drive moral progress.

Our second hypothesis is that exclusivist moralities will develop even where harsh conditions no longer exist, so long as people come to believe these conditions exist. This is where the role of social moral epistemology—or the study of the social promulgation of morally relevant beliefs—comes into play. In the adaptive plasticity model of moral development, what matters is that certain cues that are heuristically indicative of out-group threat are detected by the developing system. Such cues need not be

veridical in any given case. Thus inclusivist moralities can be dismantled not only if powerful elites engender actual environmental conditions that trigger exclusivist response but also if they create social epistemic environments in which people come to believe that such harsh conditions exist.

17.3 Applying the Model to Biomedical Moral Enhancement

17.3.1 From Evolutionary Facts to Psychological Inferences

Let us return to the question with which we began: how much moral progress is possible without the biomedical enhancement of human moral capacities? One might query whether evolutionary accounts of morality are useful in answering this question, given that the extent of human moral malleability turns on the nature of the proximate rather than the distal (etiological) causes of morality. Put somewhat more technically, synchronic properties 'screen off' diachronic properties of human moral psychology in relation to its alterability.

How, then, are evoliberals justified in inferring from etiological accounts of morality that there are severe restrictions on moral inclusivity? The answer is that although synchronic properties exhaust the facts that ultimately bear on the question of human moral plasticity, the synchronic properties are precisely what are at issue in these discussions. Where the causal structure of human moral psychology is opaque, evolutionary accounts can permit inferences about the nature of that structure and what it implies for alterability.

This raises the question: what epistemic role does evolutionary theorizing play in the evoliberal logic? Evolutionary theory appears to bridge a crucial implicit step in Persson and Savulescu's argument (2012), which we quoted at length in Section 17.2.2. This step involves moving from a premise about human moral psychology being adapted for small-group living with rudimentary technology to the claim that social and technological circumstances have changed radically while human moral psychology has remained fundamentally the same. If the latter partial premise concerning the fixed nature of human moral psychology could be established independently of evolutionary history, then the claims about evolutionary history would do no logical work in the argument, given the screening-off relation described above. Thus it must be that *facts about adaptation are taken to warrant the inference of psychological rigidity*, such that non-biomedical interventions are unlikely to succeed, which is then taken to warrant the conclusion that BME will be crucial for major moral progress.

The trouble is that the inference from innate adaptation to developmental rigidity is not warranted, and, without this inference, Persson and Savulescu's argument for the vital importance of BME founders. The concept of innateness as it applies to cognitive

psychological development is problematic, in part because of its pretheoretical associations with adaptive design, developmental rigidity, and species natures (Griffiths 2002). The fact that something is an instance of adaptive design does not imply that it is developmentally rigid, or insensitive to environmental inputs: developmental rigidity is not a necessary component of natural selection explanations. Equally fallacious is the inference from the fact that a trait is shared by all normal members of a given reference class of a given species to the conclusion that the trait is developmentally rigid. Whether a given trait is universal, whether it reflects adaptive design, whether it is biological in origin, and whether it is developmentally rigid are all contingent severable questions consistent with any configuration of answers. Likewise, the fact that a similar character state (e.g. a 'sense of fairness') is present in humans and non-human animals (e.g. other primates) does not, contra Persson and Savulescu (2008, p. 168, 2014, p. 107), imply that the trait is ancestral to both lineages, that it is genetically transmitted in humans, that it is culturally unalterable, or that it is even properly described as the same trait (Powell and Shea 2014).

Like Persson and Savulescu, we too appeal to an evolutionary account of exclusivist morality in order to draw inferences about the synchronic properties of human moral psychology. But the evolutionary model we propose allows for great plasticity in the range of moralities that can develop across plausible developmental environments. The upshot of this model is that efforts to advance and sustain moral progress in the form of inclusivity only go against the evolutionary grain under certain conditions, and these conditions are within our epistemic power to identify, and within our practical abilities to modify. Our thesis, therefore, is not that exclusivist morality is a predisposition that can be overcome by cultural innovations and moral education, but rather that the exclusivist predisposition is itself contingent on the presence of certain conditions that are culturally modifiable, calling into question the need for a systematic programme of BME.

17.3.2 Inclusivist Morality and Moral Catastrophe: A Tale of Two Solutions

As we noted in Section 17.2.2, evoliberals propose BME as an antidote to potential moral catastrophes, such as nuclear terrorism, genocide, and climate change. These pressing moral problems stem largely, in their account, from two factors: the rapid proliferation of powerful new technologies, on the one hand, and fixed evolutionary constraints on the human capacity for other-regard, on the other. The inclusivist dimensions of these problems lead evoliberals, given their background evolutionary assumptions, to the conclusion that BME will be a necessary part of the solution, whereas they lead us, given the model we have proposed conjoined with certain empirical data, to the opposite conclusion.

We are sceptical about the claim that BME will necessarily, or even likely, play a key role in solving the above moral problems. This reason for our scepticism is that most moral catastrophes are rooted in intergroup conflict, and there is little evidence to

suggest that BME will be capable in the reasonably near future of reducing the incidence and intensity of intergroup conflicts, whereas there is ample evidence to suggest that cultural innovations stand a far better chance of doing so. War—prehistorical and historical, symmetrical and asymmetrical, tribal and colonial—has nearly always been waged between racial, ethnonational, and religious groups (Keeley 1996), sometimes resulting in ethnic cleansing or genocide. Not many people are queuing up to fight and die for the United Nations—yet there is no shortage of heroes when it comes to defending one's enthnonational compatriots or religious groups, particularly in the face of perceived existential threat. Such conflicts are often facilitated by social moral epistemologies that exclude out-group members from the moral community or assign them a relatively low moral status (e.g. dehumanization tactics), thereby 'justifying' their persecution or annihilation. As we've seen, defects in social moral epistemologies are due in part to practices that facilitate the dissemination of false empirical beliefs, particularly under conditions of intergroup conflict, posing structural epistemic problems that are not likely to be ameliorated through biomedical intervention.

Wars are also far more likely to occur when at least one of the states involved is an autocracy, military junta, or monarchy, whereas war between democracies is virtually non-existent. And although democracies may wage war as often as any other type of state, the wars they do wage are significantly less severe than those waged by non-democratic states (Rummel 1995). Although the causal-theoretical basis of this robustly evidenced 'democratic peace' is hotly contested, it retains a near law-like status in international relations. Intergroup conflicts are also more likely to occur, and to occur in more severe forms, in the absence of institutions at the international level to ensure that the motives for going to war are legitimate and that the methods used to fight it are just. Indeed, the rule of law and constitutional democratic governance are social technologies designed for the express purpose of encouraging inclusivist behaviour. All of this gives us good reason to think that there are effective institutional solutions to the problems of intergroup conflict that drive many of the moral catastrophes that rightfully worry evoliberals. Furthermore, these institutional solutions instantiate inclusivist moral norms. For instance, democracy in its contemporary forms is premised on the principle that all people are entitled to participate in the political processes of their society. Likewise, in the case of institutions for international security, both just war norms and the humanitarian law of war presuppose universalizable judgements about war acts and apply the same standards to all parties.

Consider another major evoliberal concern: the impending moral disaster of climate change. Though not immediately apparent, the problem of climate change also has inclusivist moral dimensions, in at least two respects. First, there is empirical evidence to suggest that environmental degradation wrought by the activities of wealthy nations, in any plausible climate change scenario, will fall disproportionately on the world's worse-off populations both within and between nations, mainly because poorer people tend to live at the higher temperatures of lower latitudes (Mendelsohn, Dinar, and Williams 2006). Consequently, the greatest harms of climate change are likely to be

morally discounted by the comparably well-off countries and individuals that disproportionately produce them, unless something is done to ensure a more inclusivist response on the part of the better off. Second, on the evolutionary model we've proposed, climate change will significantly exacerbate intergroup conflict because it will shrink available resources and create harsher developmental conditions in affected parts of the world—giving rise to increasingly exclusivist moralities.

Third, we tend to discount the interests of future generations in deciding how we will interact with the environment. If we are to honour our moral commitments to future generations, our moral circle must expand to include not only strangers but also persons who are not yet in existence—in other words, our morality must become even more inclusive than it currently is, or than it even would be were morally arbitrary group-based discrimination against existing persons and non-human animals completely eliminated. Bringing future persons into the moral community would require a further expansion of our capacity for moral inclusiveness, which, according to evoliberals, is at or near its evolutionary limits.

In fact, the growing recognition that we have moral obligations to future persons is an excellent illustration of our commitment to a subject-centred morality (see Section 17.2.3). Thanks to the arrow of time, future generations can neither benefit us nor bite us back; persons of sufficiently distant future generations have no strategic capacities vis-à-vis contemporary people. Nor do contemporary people have sufficiently strong kin relations to distant future generations. Our moral commitments to distant future persons must therefore be grounded in a non-strategic, non-group-based conception of moral status. As we have seen, there has been a dramatic expansion of inclusivist morality over the last couple of centuries, and it is thus not much of a stretch to think that our moral circle could expand yet further to include anonymous individuals who will come to exist long after all existing people are gone. Such a norm has in fact spread quite rapidly over the last decade, culminating in Pope Francis' recent encyclical on climate change.

Difficulties in responding effectively to climate change stem not only from the power of self-interest and the limits of moral inclusivity, but also from flaws in social moral epistemic practices. Much of the political opposition to meaningful action on climate change in the US, for example, stems not from a failure of other-regard but from false empirical beliefs—namely, beliefs that that the evidence for anthropogenic climate change is non-existent or ambiguous, or that it is the result of liberal propaganda. This moral epistemic deficit can be attributed in part to an inability to identify appropriate expertise, which in the case of certain evangelical communities in the US translates into, and is motivated by, an unwarranted scepticism about claims emanating from the scientific community. Much of this scepticism seems to be enmeshed in a web of morally exclusivist beliefs, with scientific communities being perceived as a liberal threat to in-group identity. These social epistemic obstacles to progress on climate change cannot be ameliorated through BME-based interventions. It is simply not

credible to suppose that any genetic or pharmacological intervention could change complex webs of belief and patterns of epistemic deference. Likewise, only multilateral agreements and effective mechanisms of institutional enforcement can solve international collective action problems in relation to climate change in a timely fashion.

So far as we can tell, BME offers no promising ways of moderating the in-group/out-group psychological dynamics that engender most major moral catastrophes, from war, terrorism, and genocide to climate change and environmental degradation. As Persson and Savulescu acknowledge (2012, 2014), the pro-social effects of potential BMEs, such as increases in hormones (such as oxytocin) or other factors that enhance empathy, can vanish when kin relations and intergroup psychology are implicated (see de Dreu et al. 2010; Pinker 2011, ch. 9). Empathy refers to the combination of 'perspective taking' and experiencing vicarious emotions for others that are broadly in line with the emotions that other individuals are experiencing. Although empathy has been shown to mediate altruism, the problem is that, as Jesse Prinz puts it, 'empathy is ineluctably local' (2011, p. 228). Empathy can lead to pro-social behaviour when experienced specifically for stigmatized out-groups (Batson and Ahmad 2009), but enhanced empathy as a generalized capacity can exacerbate negative intergroup attitudes when it is not specifically directed towards out-groups, such as in competitive intergroup environments. In such cases, BMEs can accentuate exclusivist moral response, strengthening positive attitudes and behaviours towards the in-group while intensifying negatively valenced attitudes and behaviours towards out-groups.

In addition, because empathy is tightly bound to partiality, it can lead to a range of poor moral decision-making, such as unjustly favouring some individuals with whom we contingently empathize over other individuals with whom, contingently, we do not, or favouring the lives of concrete individuals over 'statistical' lives. Enhancing the biological underpinnings of pro-sociality can therefore make moral decision-making worse when group identity, locality, and concreteness are at stake—the very features of moral decision-making that Persson and Savulescu propose BME to counteract in the first place.

Thus it appears that many of the great moral problems we face stem not from a dearth of empathy per se but rather from the fact that the adequate stores of existing empathy are easily manipulated and misdirected in the service of intergroup conflict and local spheres of concern. Prinz concludes that the most effective way of promoting the general moral point of view (on which, incidentally, much of the evoliberal normative argument for BME rests) may require the eradication of empathy (2011, p. 228) in favour of a less parochial and less vicariously emotional 'concern' for others. However, 'concern' may not be a useful target for BME either, since it is only generated after an event has already been appraised as wrong, and it is the appraisal in particular that we must target if we are to drive moral progress in the direction of inclusivity. Once again, the problem boils down not to a general deficit of concern but rather to the fact that concern is not directed in the right ways (e.g. felt in response to the disregard of the

interests of out-group members). It is the parochiality of empathy and concern that should be the *Schwerpunkt* of our moral enhancement efforts in the struggle to stave off moral catastrophes, and BME as it has thus far been proposed fails to engage at this critical locus of the battle.

If in-group bias has biological roots, this suggests that, in theory, there may be biomedical interventions that could ameliorate exclusivist response. However, as the above discussion suggests, this does not imply that such interventions could be carried out without significant unintended costs. If parochialism was a necessary condition for the evolvability of human altruism, as the prevailing evolutionary explanation of morality would suggest, then we might expect altruism and parochialism to be mediated by common proximate causes in human psychological development (Bowles and Gintis 2013). Indeed, this is precisely what is suggested by studies showing that oxytocin and empathy accentuate exclusivist moral response.

Enhancing parochial altruism may thus amount to sharpening both sides of a double-edged sword: by strengthening the biological (hormonal, genetic, etc.) basis of altruism, we may unavoidably exacerbate antisocial attitudes and behaviours towards out-group members due to the causal developmental dependence of these phenomena. If this is the case, then we must look to avenues for enhancing moral motivations, or ways of ensuring that people act as if they are so motivated, that are not causally constrained in this way.

In sum, the driving ideas behind the evoliberal argument are a) that we are likely to discover biomedical interventions that strengthen pro-social attitudes and behaviours towards strangers and out-groups, b) that these interventions could be carried out without unacceptable or self-defeating costs that result from the developmental interconnectedness of altruism and parochialism, c) that these interventions could be implemented on a sufficiently massive scale, with entire democratic, autocratic and theocratic nations, and subversive terrorist organizations, incentivized to subject their citizenry or members to biomedical interventions, and d) that these interventions could be implemented with sufficient rapidity to address imminent catastrophic threats. Each of these points seems dubious.

In contrast, the model of evolutionary moral development that we have outlined suggests a number of concrete cultural avenues for addressing the gravest moral problems that we face. In addition to basic moral education, such as teaching individuals to resist their natural proclivity towards essentialistic classifications of human groups, concerted efforts must be made to ameliorate environmental conditions that mimic the dangerous conditions of early human evolution. For it is only in these 'luxurious' circumstances that inclusivist morality can take root and expand.

17.3.3 A Nontraditional Approach to Traditional Moral Enhancement

The effort to modify conditions that trigger exclusivist moral response involves several, interrelated components. The first involves creating an environment of physical and economic security, both internationally and in microenvironments within otherwise

secure nations. By fostering economic productivity and social surpluses (e.g. by instituting markets, along with effective property rights and the rule of law), encouraging the genuine democratization of political institutions, and creating institutions that allow for mutually beneficial intergroup cooperation and the peaceful resolution of intergroup conflicts (as now exist at both domestic and to a lesser but still meaningful extent international levels), we can reduce and ultimately eliminate many of the ancient trigger conditions that cue the development of exclusivist response. None of these cultural innovations require intervening at the level of individual moral capacities, as BME promises to do.

Second, by significantly reducing the incidence of infectious disease, we can lessen the effects of yet another major out-group threat cue type—namely, signs of parasite stress, which have been shown experimentally to have antisocial priming effects, leading to increased xenophobia and ethnocentricism (reviewed in Fincher and Thornhill 2011). Although not all disease and disability is infectious and thus not all disease and disability indicates the existence of parasite threat, the evolution of adaptive plasticity is a heuristic process that is epistemically incapable of finely discriminating between disease cues based on their etiology. Thus a broad strokes approach to reducing general rates of disease through sanitation, vaccination, and broader public and private healthcare initiatives is likely to have a significant attenuating effect on the development of exclusivist moralities. Although such interventions are biomedical in nature, they do not fall under the rubric of BME as standardly conceived, insofar as the latter concerns the direct modification of specifically moral capacities. These conventional biomedical interventions are likely to be more efficacious, cost-effective, and logistically feasible than BME programmes when it comes to modulating exclusivist moral tendencies.

Third, we must ensure that inclusivist cultural innovations, such as the protections afforded by the recognition and institutionalization of human rights in the domestic and international spheres, are not dismantled by social epistemic practices that are designed to engender perceptions of out-group threat conditions, including, pre-eminently, propaganda designed to evoke racial or enthnonational violence. Enhancing social moral epistemic practices so as to increase their reliability in producing correct morally relevant beliefs cannot be accomplished biomedically, let alone through direct BME-based interventions that target the moral capacities of individual people. In focusing on enhancing individual cognitive and affective capabilities, BME has tended to overlook the fact that knowledge crucial for good moral decision-making is the product of social practices, and thus that the potential for moral progress often turns on the epistemic virtues of those social practices.

All of the above trigger conditions can causally feed back into one another. For example, the lack of a stable healthcare infrastructure can lead to an environment replete with cues of parasite stress and can contribute to economic distress. Both of these conditions in turn can engender exclusivist moral tendencies that drive intergroup conflicts, which are only exacerbated by the lack of institutions that protect the rights

of minority groups and provide opportunities for peaceful conflict resolution. Such conditions are conducive to the development of social moral epistemologies that reinforce and exacerbate group conflicts through the dissemination of false empirical beliefs about out-groups and other circumstances indicative of out-group threat. These faulty moral epistemologies in turn can lead to the institutionalized mistreatment of out-groups (including their exclusion from employment, ghettoization, etc.), the results of which seem to confirm the faulty epistemologies that were used to justify their mistreatment in the first place (e.g. they are disease-ridden, irrational animals prone to criminality and deceit, etc.), resulting in the solidification of an exclusivist morality. On the other hand, cultural innovations that foster inclusivity, especially by extending human rights and basic civil liberties to all people, can help break these vicious causal and epistemic spirals.

An important difference between the approach we advocate and the evoliberal emphasis on BME is that our approach targets population-level statistical effects on the social development of morality, whereas the evoliberal approach aims for a direct, immediate impact on individual moral development. This difference in our respective approaches bespeaks an important philosophical difference in our respective conceptions of morality. In our view, morality (as defined in Section 17.2.1) is not an epiphenomenon that supervenes on the aggregate of individual moral capacities and judgements constrained by evolutionary history. It is a dynamic social phenomenon that causally feeds back into the processes of individual moral development that produce it, which in turn serve as causal inputs into the social evolution of moral systems, and so on.

This feedback process resembles the phenomenon of niche construction in biology, wherein organismic adaptations shape the ecological environments in which those same organisms and adaptations continue to develop and evolve. In social moral evolution, this feedback process can drive both moral progression and moral regression, depending on the direction of the changes to the initial conditions of the system (as described above). BME proponents, on the other hand, by focusing on individual moral response generated by a biologically rigid human moral nature, tend to overlook this dynamic, biocultural causal structure of human morality. In essence, we believe that the evolutionary mismatch problem, as it has framed the BME debate (Section 17.1), is misconceived. Human moral psychology is not a rigid square peg trying to fit in the round hole of modern human ecology. Rather, the shape of the peg is determined in part by the shape of the hole.

17.3.4 Two Kinds of Moral Progress and the Role of Institutions

This brings us to an important point that tends to get lost in discussions of BME, which is that advocates of BME tend to underplay the pivotal role of institutions in driving moral progress.

The threat of punishment exerts a much stronger influence on pro-social behaviour than oxytocin or other BME variables have been shown to do. This is confirmed by

laboratory studies involving economic games, as well as in the real world where institutions of contract, property, criminal, and tort law and their associated enforcement mechanisms exert a far more profound and positive influence on mutually beneficial cooperation, promise-keeping, non-exploitation, and non-aggression than BMEs are ever likely to do.

This raises an important ambiguity in the term 'moral enhancement' and how it relates to moral progress. Moral enhancement might refer to improvements in moral motivations, such as the ability to feel compassion or love for others within or beyond one's social group. Let us call this first, internal type of moral enhancement ME1. Alternatively, 'moral enhancement' might refer to interventions external to the agent that generate behaviour that is in accordance with correct moral principles, irrespective of its motivations. Let us call this second, external type of moral enhancement ME2. As we have seen, ME2 can occur under the right institutional environments, even if altruistic or sympathetic motivations remain unchanged. Institutions can encourage people to act as if, for example, foreigners or domestic minorities are of equal moral worth, even if they are incapable of loving or empathizing with them. By the same token, even if people have laudable moral motivations, social institutional circumstances may be such that they are incapable of acting on these motivations (as was the case, perhaps, for many German citizens under the Third Reich).

Steven Pinker (2011) argues in his recent magnum opus that homicide rates have plummeted over the last few centuries due to institutional changes that brought about much better compliance with moral norms against killing. This progressive trend in moral behaviour may have occurred even if people, so far as their moral attitudes and motivations are concerned, did not improve. Furthermore, according to the adaptive plasticity model of evolutionary moral development that we have proposed, the reduction of intergroup conflict brought about by ME2 may have psychological spillover effects into ME1. One possibility, for example, is that ME1 becomes developmentally possible only when cues of out-group threat have been sufficiently diminished by ME2. Indeed, there is a great deal of evidence that the creation of the state, and with it a monopoly on the use of force, massively reduced homicide rates and allowed for more peaceful interactions between individuals (see e.g. Elias 2000; Pinker 2011), providing conditions under which ME1 could begin to change. Recognizing the conditional evolutionary limitations on ME1, and deliberately attempting to work around them by engaging in ME2, is itself a form of moral improvement.

Even if evolved constraints on other-regard and moral imagination were insurmountable, this need not entail severe constraints on moral behaviour since the latter could still be dramatically shaped by institutions. Persson and Savulescu make much of the fact that there has been little moral progress in over 2500 years since the first great teachers of morality (2014, p. 106). Pinker's review and analysis (2011), however, belies the claim that moral progress has been extremely limited, and in another paper

(Buchanan and Powell, unpublished) we identify several distinct types of moral progress that have occurred. Moreover, we can begin to make sense of the fact that most moral progress has taken place roughly over the last 250 years, despite the ancient history of moral philosophizing, once our focus shifts away from moral reasoning and towards the conditions and institutions that are necessary for moral scrutiny to flourish and become socially efficacious, which (as we suggested earlier) have arisen only very recently in human history.

It is true that cultural moral innovations that foster inclusivist morality are unlikely to stop the disturbed lone wolf or disaffected terrorist group hell bent on revenge, creating chaos or resurrecting fascism, but then neither is a systematic programme of BME, even if it were ultimately shown to be effective. The notion that we could morally enhance the overwhelming majority of the human population by instituting a global BME programme, and that these BMEs would significantly reduce the rates of random acts of violence, let alone atrocities committed during intergroup conflict, is a hard pill to swallow. It strains the imagination to envision the level of global political control that would be required to ensure that every human being was psychologically re-engineered, especially since terrorists and genocidaires would not be likely to queue up for treatment.

17.4 From an Optimistic Induction to an Evomoderate Conclusion

In arguing that a non-traditional approach to traditional moral enhancement is far more promising than BME, we do not mean to understate the daunting nature of the task that lies before us. Establishing lasting economic, political, healthcare, and security infrastructures, as well as international institutions that successfully mediate interstate conflict, solve collective action problems, and help ensure that basic human rights are respected within states, is clearly a monumental undertaking. Nevertheless, in light of what we are coming to know about morality and its evolution, we believe that cultural innovation is the best way to prevent moral catastrophes such as genocide, nuclear war, terrorism, and climate change, and to ensure that our moral circle continues to expand.

This naturalized history of moral progress leaves us cautiously optimistic that, despite its evolutionary psychological baggage, humanity has the cultural resources and can develop the necessary social technologies to shoulder the great moral burdens it now faces. We therefore disagree with Persson and Savulescu's claim that 'Liberal democracies cannot overcome these problems by developing novel technology' (2012, p. 168), if 'technology' is construed broadly to include legal institutions, markets, public health infrastructures, and so on. The notion that we have reached the outer bounds of human moral inclusivity, or that further inclusivist expansions cannot keep pace with technological development—and hence, that we need to embark on a systematic

programme of BME—is contrary to historical evidence and inconsistent with the dynamic model of evolutionary development we have put forward here. We have in effect staked out what might be called an 'evomoderate' position—one that acknowledges evolved constraints on human dispositions but also recognizes the great flexibility of human moral thinking and practice that allows these constraints to be identified and overcome without any changes to the 'fundamental' biological constitution of human beings.

Research on BME is still in its infancy, however, and we think it is reasonable to view biomedical intervention as one potential instrument in our diverse moral enhancement toolkit. Like many ethicists writing on this topic, we see no in-principle objection to using biomedical technologies in conjunction with cultural modes of moral enhancement to bring moral motivations and behaviours in line with the norms we have come to endorse. Nevertheless, the foregoing analysis leads us to the pessimistic conclusion that BME is unlikely to play a necessary, or even major, role in the future of moral progress or in solving the greatest moral dilemmas of the coming centuries. We agree with the evoliberal headline that there is an 'urgent need to enhance the moral character of humanity' (Persson and Savulescu 2008), but we do not think that BME is an effective and plausible means by which to do so.

Acknowledgements

Russell Powell is grateful to Templeton Foundation Grant # 43160 for support of this research.

References

Agar, N. (2013). *Truly Human Enhancement: A Philosophical Defense of Limits*. Cambridge, MA: MIT Press.

Alexander, R. D. (1987). *The Biology of Moral Systems*. New York: Aldine de Gruyter.

Asma, S. (2012). *Against Fairness*. Chicago: University of Chicago Press.

Batson, C. D. and Ahmad, N. Y. (2009). 'Using Empathy to Improve Intergroup Attitudes and Relations', *Social Issue and Policy Review* 3(1): 141–77.

Boehm, C. (2001). *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Cambridge, MA: Harvard University Press.

Buchanan, A. (1990). 'Justice as Reciprocity versus Subject-Centered Justice', *Philosophy and Public Affairs* 19: 227–52.

Buchanan, A. (2011). *Beyond Humanity? The Ethics of Biomedical Enhancement*. Oxford: Oxford University Press.

Buchanan, A. and Powell, R. (2015). 'The Limits of Evolutionary Explanations of Morality and their Implications for Moral Progress', *Ethics* 126(1): 37–67.

Buchanan, A. and Powell, R. (forthcoming), 'Toward a Naturalistic Theory of Moral Progress', *Ethics*.

Buchanan, A. and Powell, R. (unpublished manuscript), 'What is Moral Progress?'.

Campbell, R. and Kumar, V. (2012). 'Moral Reasoning on the Ground', *Ethics* 122(2): 273–312.

De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., Van Kleef, G. A., Baas, M., Ten Velden, F. S., Van Dijk, E., and Feith, S. W. W. (2010). 'The Neuropeptide Oxytocin Regulates Parochial Altruism in Intergroup Conflict Among Humans', *Science* 328(5984): 1408–11.

DeScioli, P. and Kurzban, R. (2013). 'A Solution to the Mysteries of Morality', *Psychological Bulletin* 139(2): 477–96.

Elias, N. (2000). *The Civilizing Process: Sociogenetic and Psychogenetic Investigations*. Oxford: Blackwell.

Fincher, C. L. and Thornhill, R. (2011). 'Parasite-Stress Promotes In-group Assortative Sociality: the Cases of Strong Family Ties and Heightened Religiosity', *Behavioral and Brain Sciences* 35(2): 61–79.

Gaertner, S. L. and Dovidio, J. F. (2000). *Reducing Intergroup Bias: The Common Ingroup Identity Model*. Philadelphia, PA: Psychology Press.

Ghiselin, M. (1974). *The Economy of Nature and the Evolution of Sex*. Berkeley, CA: University of California Press.

Bowles, S. and Gintis, H. (2013). *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton, NJ: Princeton University Press.

Goodman, R. and Jinks, D. (2013). *Socializing States: Promoting Human Rights through International Law*. Oxford: Oxford University Press.

Griffiths, P. (2002). 'What is Innateness?' *The Monist* 85(1): 70–85.

Haidt, J. (2001). 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment', *Psychological Review* 108(4): 814–34.

Haidt, J. (2012). *The Righteous Mind*. New York: Pantheon.

Harris, J. (2010). *Enhancing Evolution: the Ethical Case for Making Better People*. Princeton: Princeton University Press.

Keeley, L. (1996). *War Before Civilization*. Oxford: Oxford University Press.

Lazari-Radek, K. D. and Singer, P. (2012). 'The Objectivity of Ethics and the Unity of Practical Reason', *Ethics* 123: 9–31.

Mendelsohn, R., Dinar, A., and Williams, L. (2006). 'The Distributional Impact of Climate Change on Rich and Poor Countries', *Environment and Development Economics* 11: 159–78.

Moody-Adams, M. (1999). 'The Idea of Moral Progress', *Metaphilosophy* 30(3): 168–85.

Persson, I. and Savulescu, J. (2008). 'The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity', *Journal of Applied Philosophy* 25(3): 162–7.

Persson, I. and Savulescu, J. (2012). 'Moral Enhancement, Freedom and the God Machine', *Monist* 95(3): 399–421.

Persson, I. and Savulescu, J. (2014). *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press.

Pinker, S. (2011). *The Better Angels of Our Nature*. New York: Viking.

Powell, R. and Shea, N. (2014). 'Homology Across Inheritance Systems', *Biology and Philosophy* 29(6): 781–806.

Prinz, J. (2011). 'Against Empathy', *The Southern Journal of Philosophy* 49: 214–33.

Richerson, P. and Boyd, R. (2005). *Not in Our Genes*. Chicago: University of Chicago Press.

Rummel, R. J. (1995). 'Democracies are Less Warlike than Other Regimes', *European Journal of International Relations* 1(4): 457–47.

Singer, P. (2011). *Expanding the Moral Circle: Ethics, Evolution, and Moral Progress*. Princeton, NJ: Princeton University Press.

Wrangham, R. and Peterson, D. (1996). *Demonic Males: Apes and the Origins of Human Violence*. New York: Houghton Mifflin.

Index